

GenCore version 5.1.6
 Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Title: US-09-084-691B-206
 Perfect score: 1045
 Sequence: MSTLKPDKTKRTRNRRPT.....CSFSIFLALLSCUTTPASA 191

Run on: May 12, 2004, 09:08:05 ; Search time 57 Seconds
 (without alignments)
 946.782 Million cell updates/sec

Scoring table: BLOSUM62
 Gapop 10.0 , Gapext 0.5
 Seq ed: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107
 Minimum DB seq length: 0
 Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
 Maximum Match 100%
 Listing first 90 summaries

Database : A_Genesed_29Jan04:
 1: genesedGP1980s:
 2: genesedGP1990s:
 3: genesedGP2000s:
 4: genesedGP2001s:
 5: genesedGP2002s:
 6: genesedGP2003as:
 7: genesedGP2003bs:
 8: genesedGP2004s:
 9: genesedGP2004s:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result | Score | Query | Match | Length | DB | ID | Description |
|--------|-------|-------|-------|--------|----------|----|-------------------------|
| 1: | 1045 | 100.0 | 191 | 2 | AAR92987 | | Aar92987 Hepatitis |
| 2: | 996 | 95.3 | 191 | 2 | AAR92972 | | Aar92972 Hepatitis |
| 3: | 982 | 94.0 | 191 | 2 | AAR92974 | | Aar92974 Hepatitis |
| 4: | 982 | 94.0 | 191 | 2 | AAR92973 | | Aar92973 Hepatitis |
| 5: | 981 | 93.9 | 191 | 2 | AAR92953 | | Aar92953 Hepatitis |
| 6: | 980 | 93.8 | 191 | 2 | AAR92978 | | Aar92978 Hepatitis |
| 7: | 978 | 93.6 | 191 | 2 | AAR92976 | | Aar92976 Hepatitis |
| 8: | 977 | 93.5 | 191 | 2 | AAR92977 | | Aar92977 Hepatitis |
| 9: | 977 | 93.5 | 191 | 2 | AAR74048 | | Aar74048 Synthetic |
| 10: | 977 | 93.5 | 196 | 2 | AAR74047 | | Aar74047 Synthetic |
| 11: | 977 | 93.5 | 319 | 2 | AAR96546 | | Aar96546 Hepatitis |
| 12: | 977 | 93.5 | 326 | 2 | AAR22137 | | Aar22137 HCV-HCg9 |
| 13: | 977 | 93.5 | 733 | 2 | AAR38278 | | Aar38278 NANB hepatitis |
| 14: | 977 | 93.5 | 2894 | 2 | AAR24440 | | Aar24440 Composite |
| 15: | 977 | 93.5 | 2894 | 2 | AAR70230 | | Aar70230 Composite |
| 16: | 977 | 93.5 | 3011 | 2 | AAR66995 | | Aar66995 Hepatitis |
| 17: | 976 | 93.4 | 191 | 2 | AAR44010 | | Aar44010 Hepatitis |
| 18: | 976 | 93.4 | 191 | 2 | AAR92938 | | Aar92938 Hepatitis |
| 19: | 976 | 93.4 | 191 | 2 | AAR92941 | | Aar92941 Hepatitis |
| 20: | 976 | 93.4 | 191 | 2 | AAR92937 | | Aar92937 Hepatitis |
| 21: | 976 | 93.4 | 191 | 2 | AAR92939 | | Aar92939 Hepatitis |
| 22: | 976 | 93.4 | 191 | 2 | AAR92940 | | Aar92940 Hepatitis |
| 23: | 976 | 93.4 | 191 | 3 | AAY9411 | | Aay9411 Human hepatitis |
| 24: | 976 | 93.4 | 502 | 2 | AAR67591 | | Aar67591 Hepatitis |
| 25: | 976 | 93.4 | 2 | | AAR79222 | | Aar79222 phcv141-e |

ALIGNMENTS

RESULT 1
 ID AAR92987 standard; protein: 191 AA.

XX
AC AAR92987;
XX 02-OCT-1996 (first entry)
XX DE Hepatitis C virus isolate HK2 core protein.
XX KW HCV; E1; envelope 1; core protein; HCV genotyping; antibody; vaccine;
XX hepatitis.
XX OS Hepatitis C virus.
XX PN WO9605315-A2..
XX PD 22-FEB-1996.
XX PP 15-AUG-1995; 95WO-US010398.
XX PR 15-AUG-1994; 94US-00290665.
XX F (USSH) US SEC DEPT HEALTH.
XX PI Bukh J, Miller RH, Purcell RH;
XX DR WPI; 1996-139709/14.
XX N-PSDB; AAT16661.
XX DNA and amino acid sequence of HCV envelope 1 and core proteins - used to
PT determine HCV genotype and as vaccines against HCV infection.
XX Claim 4; Page 223; 340pp; English.
XX AR92936-R32987 are HCV core proteins derived from 52 different HCV
CC isolates. Isolated cDNA sequences are used for the prodn. of primers
CC useful for detecting the presence of HCV in a sample, the primers are
CC also useful for HCV genotyping. Proteins encoded by the cDNAs can be used
CC in vaccines for immunising against HCV infection. The proteins may also
CC be used to detect antibodies against HCV in serum, saliva, lymphocytes or
CC other mononuclear cells. The antibodies may be used in the prevention of
CC HCV infection.
XX Sequence 191 AA;
Query Match 100.0%; Score 1045; DB 2; length 191;
Best Local Similarity 100.0%; Pred. No. 2.4e-90;
Matches 191; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MSTLPKQRKTAKNTNRPTDVKPGQQIVGGVYLPRRGPRPLGRYTRAKTSERSOPRG 60
DQ 1 MSTLPKQRKTAKNTNRPTDVKPGQQIVGGVYLPRRGPRPLGRYTRAKTSERSOPRG 60
QY 61 RQQPPIKARQPRGHRWAPQGPWPLYNGEGCWAGWILSPRSRPRPDRRSRLG 120
DQ 61 RQQPPIKARQPRGHRWAPQGPWPLYNGEGCWAGWILSPRSRPRPDRRSRLG 120
Db 121 KVIDTLCGFADLIMGYIPVGAPLGVAALAHGVRAVEDGINYATGNLPGCSFSIFL 180
QY 121 KVIDTLCGFADLIMGYIPVGAPLGVAALAHGVRAVEDGINYATGNLPGCSFSIFL 180
DQ 121 KVIDTLCGFADLIMGYIPVGAPLGVAALAHGVRAVEDGINYATGNLPGCSFSIFL 180
QY 181 LLSCLTPASA 191
Db 181 LLSCLTPASA 191
RESULT 2
ID AAR92972
ID AAR92972 standard; protein; 191 AA.
AC AAR92972;
XX 02-OCT-1995 (first entry)
XX DE Hepatitis C virus isolate Z4 core protein.
XX KW HCV; E1; envelope 1; core protein; HCV genotyping; antibody; vaccine;
XX hepatitis.
XX OS Hepatitis C virus.
XX PN WO9605315-A2..
XX PD 22-FEB-1996.
XX PP 15-AUG-1995; 95WO-US010398.
XX PR 15-AUG-1994; 94US-00290665.
XX PA (USSH) US SEC DEPT HEALTH.
XX PI Bukh J, Miller RH, Purcell RH;
XX DR WPI; 1996-139709/14.
XX N-PSDB; AAT16646.
XX DNA and amino acid sequence of HCV envelope 1 and core proteins - used to
PT determine HCV genotype and as vaccines against HCV infection.
XX PS Claim 4; Page 210-211; 340pp; English.
XX AR92936-R32987 are HCV core proteins derived from 52 different HCV
CC isolates. Isolated cDNA sequences are used for the prodn. of primers
CC useful for detecting the presence of HCV in a sample, the primers are
CC also useful for HCV genotyping. Proteins encoded by the cDNAs can be used
CC in vaccines for immunising against HCV infection. The proteins may also
CC be used to detect antibodies against HCV in serum, saliva, lymphocytes or
CC other mononuclear cells. The antibodies may be used in the prevention of
CC HCV infection.
XX Sequence 191 AA;
Query Match 95.3%; Score 996; DB 2; length 191;
Best Local Similarity 94.8%; Pred. No. 9.7e-86;
Matches 181; Conservative 4; Mismatches 6; Indels 0; Gaps 0;
QY 1 MSTLPKQRKTAKNTNRPTDVKPGQQIVGGVYLPRRGPRPLGRYTRAKTSERSOPRG 60
DQ 1 MSTLPKQRKTAKNTNRPTDVKPGQQIVGGVYLPRRGPRPLGRYTRAKTSERSOPRG 60
QY 61 RQQPPIKARQPRGHRWAPQGPWPLYNGEGCWAGWILSPRSRPRPDRRSRLG 120
DQ 61 RQQPPIKARQPRGHRWAPQGPWPLYNGEGCWAGWILSPRSRPRPDRRSRLG 120
Db 121 KVIDTLCGFADLIMGYIPVGAPLGVAALAHGVRAVEDGINYATGNLPGCSFSIFL 180
QY 121 KVIDTLCGFADLIMGYIPVGAPLGVAALAHGVRAVEDGINYATGNLPGCSFSIFL 180
DQ 121 KVIDTLCGFADLIMGYIPVGAPLGVAALAHGVRAVEDGINYATGNLPGCSFSIFL 180
QY 181 LLSCLTPASA 191
Db 181 LLSCLTPASA 191
RESULT 3
ID AAR92974
ID AAR92974 standard; protein; 191 AA.
AC AAR92974;
XX 02-OCT-1996 (first entry)
XX DE Hepatitis C virus isolate Z1 core protein.
XX KW HCV; E1; envelope 1; core protein; HCV genotyping; antibody; vaccine;
XX hepatitis.
XX OS Hepatitis C virus.

PN WO9605315-A2.
 XX
 DD 22-FEB-1995.
 XX
 PF 15-AUG-1995; 95WO-US010398.
 XX
 PR 15-AUG-1994; 94US-00290665.
 XX
 PA (USSH) US SEC DEPT HEALTH.
 XX
 PT Bukh, J, Miller RH, Purcell RH;
 DR
 XX
 DR N-PSDB; AATI6648.
 XX
 PT DNA and amino acid sequence of HCV envelope 1 and core proteins - used to determine HCV genotype and as vaccines against HCV infection.
 XX
 PT AR92936-R92987 are HCV core proteins derived from 52 different HCV isolates. Isolated cDNA sequences are used for the prodn. of primers useful for detecting the presence of HCV in a sample, the primers are also useful for HCV genotyping. Proteins encoded by the cDNAs can be used in vaccines for immunising against HCV infection. The proteins may also be used to detect antibodies against HCV in serum, saliva, lymphocytes or other mononuclear cells. The antibodies may be used in the prevention of HCV infection.
 XX
 SQ Sequence 191 AA;
 Query Match 94.0%; Score 982; DB 2; Length 191;
 Best Local Similarity 93.7%; Pred. No. 2e-84; Mismatches 5; Indels 0; Gaps 0;
 Matches 179; Conservative 5;
 QY 1 MSTLPKQRKTKRNTNRPTDVKFPGCGQIVGCVYLPRRGRLGRVTRKTSERSOPRG 60
 1 MSTNPKPQRKTKRNTNRPTDVKFPGCGQIVGCVYLPRRGRLGRVTRKTSERSOPRG 60
 Db 61 RROPIPKARQPOGRHAWQPGVWPLKLYNGEKGWAGMILSPGSRPHGPNPDRRSRNLG 120
 61 RROPIPKARQPOGRHAWQPGVWPLKLYNGEKGWAGMILSPGSRPHGPNPDRRSRNLG 120
 Db 61 RROPIPKARQPOGRHAWQPGVWPLKLYNGEKGWAGMILSPGSRPHGPNPDRRSRNLG 120
 61 RROPIPKARQPOGRHAWQPGVWPLKLYNGEKGWAGMILSPGSRPHGPNPDRRSRNLG 120
 QY 121 KVIDLTTCGFADLMGYTPVVGAPLGGVAAALAHGVRAIEDGINYATGNLPGCSFSIFLLA 180
 121 KVIDLTTCGFADLMGYTPVVGAPLGGVAAALAHGVRAIEDGINYATGNLPGCSFSIFLLA 180
 Db 121 KVIDLTTCGFADLMGYTPVVGAPLGGVAAALAHGVRAIEDGINYATGNLPGCSFSIFLLA 180
 121 KVIDLTTCGFADLMGYTPVVGAPLGGVAAALAHGVRAIEDGINYATGNLPGCSFSIFLLA 180
 Db 181 LLSCLTPASA 191
 181 LLSCLTPASA 191
 181 LLSCLTPASA 191
 RESULT 4
 AAR92973
 ID AAR92973 standard; protein; 191 AA.
 XX
 AC AAR92973;
 XX
 DT 02-OCT-1996 (first entry)
 XX
 DE Hepatitis C virus isolate Z8 core protein.
 XX
 KW HCV; E1; envelope 1; core protein; HCV genotyping; antibody; vaccine;
 KW hepatitis.
 XX
 OS Hepatitis C virus.
 XX
 PN WO9605315-A2.
 XX
 PD 22-FEB-1996.
 XX
 PR 15-AUG-1995; 95WO-US010398.
 XX
 PA (USSH) US SEC DEPT HEALTH.
 XX
 PT Bukh, J, Miller RH, Purcell RH;
 XX
 DR N-PSDB; AATI6647.
 XX
 PT DNA and amino acid sequence of HCV envelope 1 and core proteins - used to determine HCV genotype and as vaccines against HCV infection.
 XX
 PS Claim 4; Page 211-212; 340pp; English.
 XX
 CC AR92936-R92987 are HCV core proteins derived from 52 different HCV isolates. Isolated cDNA sequences are used for the prodn. of primers useful for detecting the presence of HCV in a sample, the primers are also useful for HCV genotyping. Proteins encoded by the cDNAs can be used in vaccines for immunising against HCV infection. The proteins may also be used to detect antibodies against HCV in serum, saliva, lymphocytes or other mononuclear cells. The antibodies may be used in the prevention of HCV infection.
 XX
 SQ Sequence 191 AA;
 Query Match 94.0%; Score 982; DB 2; Length 191;
 Best Local Similarity 93.7%; Pred. No. 2e-84; Mismatches 5; Indels 0; Gaps 0;
 Matches 179; Conservative 5;
 QY 1 MSTLPKQRKTKRNTNRPTDVKFPGCGQIVGCVYLPRRGRLGRVTRKTSERSOPRG 60
 1 MSTNPKPQRKTKRNTNRPTDVKFPGCGQIVGCVYLPRRGRLGRVTRKTSERSOPRG 60
 Db 61 RROPIPKARQPOGRHAWQPGVWPLKLYNGEKGWAGMILSPGSRPHGPNPDRRSRNLG 120
 61 RROPIPKARQPOGRHAWQPGVWPLKLYNGEKGWAGMILSPGSRPHGPNPDRRSRNLG 120
 Db 121 KVIDLTTCGFADLMGYTPVVGAPLGGVAAALAHGVRAIEDGINYATGNLPGCSFSIFLLA 180
 121 KVIDLTTCGFADLMGYTPVVGAPLGGVAAALAHGVRAIEDGINYATGNLPGCSFSIFLLA 180
 QY 181 LLSCLTPASA 191
 181 LLSCLTPASA 191
 RESULT 5
 AAR92953
 ID AAR92953
 XX
 AC AAR92953;
 XX
 DT 02-OCT-1996 (first entry)
 XX
 DE Hepatitis C virus isolate HK3 core protein.
 XX
 KW HCV; E1; envelope 1; core protein; HCV genotyping; antibody; vaccine;
 KW hepatitis.
 XX
 OS Hepatitis C virus.
 XX
 PN WO9605315-A2.
 XX
 PD 22-FEB-1996.
 XX
 PR 15-AUG-1994; 94US-00290665.
 XX
 PA (USSH) US SEC DEPT HEALTH.
 XX
 PT Bukh, J, Miller RH, Purcell RH;
 XX

DR WPI; 1996-139709/14.
 DR N-PSDB; AATI6627.
 XX
 PT DNA and amino acid sequence of HCV envelope 1 and core proteins - used to determine HCV genotype and as vaccines against HCV infection.
 XX
 PS Claim 4; Page 194-195; 340pp; English.
 XX
 XX AR92936-R92987 are HCV core proteins derived from 52 different HCV isolates. Isolated cDNA sequences are used for the prodn. of primers useful for detecting the presence of HCV in a sample, the primers are also useful for HCV genotyping. Proteins encoded by the cdnas can be used in vaccines for immunising against HCV infection. The proteins may also be used to detect antibodies against HCV in serum, saliva, lymphocytes or other mononuclear cells. The antibodies may be used in the prevention of HCV infection.

SQ Sequence 191 AA:

Query Match 93.9%; Score 981; DB 2; Length 191;
 T Local Similarity 93.7%; Pred. No. 2.5e-84; Mismatches 5; Indels 0; Gaps 0;
 Cches 179; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

QY 1 MSTLPKERRQTKRNTNRPTDKFPGSGQIVGGVYLPRGRGLGVTRAKTSERSOPRG 60
 1 MSTNPKPQKRTKRNTNRPTDKFPGSGQIVGGVYLPRGRGLGVTRAKTSERSOPRG 60

Db 61 RQPPIKARQPOQRHWAQPGVWPLXGNGEGCGWAGWLLSPRSRPWGPNDPRRSRNLG 120
 61 RQPPIKARQPOQRHWAQPGVWPLXGNGEGCGWAGWLLSPRSRPWGPNDPRRSRNLG 120

Qy 121 KVDITLTCGFADLMGYIPVVGAPLGVAARALAHGVRLLEDGVNYATGNLPGCSFSIFLA 180
 121 KVDITLTCGFADLMGYIPVVGAPLGVAARALAHGVRLLEDGVNYATGNLPGCSFSIFLA 180

Db 181 LLSCLTPASA 191
 181 LLSCLTPASA 191

RESULT 6

ID AAR92978
 ID AAR92978 standard; protein: 191 AA.
 XX
 AC AAR92978;
 XX
 DT 02-OCT-1996 (first entry)

DR Hepatitis C virus isolate DK13 core protein.
 XX
 XN HCV; El; envelope 1; core protein; HCV genotyping; antibody; vaccine; hepatitis.

XX
 OS Hepatitis C virus.

XX
 PN WO9605315-A2.
 XX
 PD 22-FEB-1995.

XX
 PP 15-AUG-1995; 95WO-US010398.

XX
 PR 15-AUG-1994; 94US-00290665.

XX
 PA (USSH) US SEC DEPT HEALTH.

XX
 PT Bukh, J., Miller RH, Purcell RH,
 XX
 DR WPI; 1996-139709/14.
 DR N-PSDB; AATI6650.

XX
 XX DNA and amino acid sequence of HCV envelope 1 and core proteins - used to determine HCV genotype and as vaccines against HCV infection.

XX
 PS Claim 4; Page 214; 340pp; English.

XX
 XX AR92936-R92987 are HCV core proteins derived from 52 different HCV isolates. Isolated cDNA sequences are used for the prodn. of primers useful for detecting the presence of HCV in a sample, the primers are also useful for HCV genotyping. Proteins encoded by the cdnas can be used

PS Claim 4; Page 215-216; 340pp; English.
 XX
 CC AAR92936-R92987 are HCV core proteins derived from 52 different HCV isolates. Isolated cDNA sequences are used for the prodn. of primers useful for detecting the presence of HCV in a sample, the primers are also useful for HCV genotyping. Proteins encoded by the cdnas can be used in vaccines for immunising against HCV infection. The proteins may also be used to detect antibodies against HCV in serum, saliva, lymphocytes or other mononuclear cells. The antibodies may be used in the prevention of HCV infection.

SQ Sequence 191 AA:

Query Match 93.8%; Score 980; DB 2; Length 191;
 T Local Similarity 93.7%; Pred. No. 3.1e-84; Mismatches 4; Indels 0; Gaps 0;
 Cches 179; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY 1 MSTLPKERRQTKRNTNRPTDKFPGSGQIVGGVYLPRGRGLGVTRAKTSERSOPRG 60
 1 MSTNPKPQKRTKRNTNRPTDKFPGSGQIVGGVYLPRGRGLGVTRAKTSERSOPRG 60

Db 61 RQPPIKARQPOQRHWAQPGVWPLXGNGEGCGWAGWLLSPRSRPWGPNDPRRSRNLG 120
 61 RQPPIKARQPOQRHWAQPGVWPLXGNGEGCGWAGWLLSPRSRPWGPNDPRRSRNLG 120

Qy 121 KVDITLTCGFADLMGYIPVVGAPLGVAARALAHGVRLLEDGVNYATGNLPGCSFSIFLA 180
 121 KVDITLTCGFADLMGYIPVVGAPLGVAARALAHGVRLLEDGVNYATGNLPGCSFSIFLA 180

Db 181 LLSCLTPASA 191
 181 LLSCLTPASA 191

RESULT 7

ID AAR92976
 ID AAR92976 standard; protein: 191 AA.
 XX
 AC AAR92976;
 XX
 DT 02-OCT-1996 (first entry)

DR Hepatitis C virus isolate Z6 core protein.
 XX
 XN HCV; El; envelope 1; core protein; HCV genotyping; antibody; vaccine; hepatitis.

XX
 OS Hepatitis C virus.

XX
 PN WO9605315-A2.
 XX
 PD 22-FEB-1996.

XX
 PP 15-AUG-1995; 95WO-US010398.

XX
 PR 15-AUG-1994; 94US-00290665.

XX
 PA (USSH) US SEC DEPT HEALTH.

XX
 PT Bukh, J., Miller RH, Purcell RH,
 XX
 DR WPI; 1996-139709/14.
 DR N-PSDB; AATI6650.

XX
 XX DNA and amino acid sequence of HCV envelope 1 and core proteins - used to determine HCV genotype and as vaccines against HCV infection.

XX
 PS Claim 4; Page 214; 340pp; English.

XX
 XX AR92936-R92987 are HCV core proteins derived from 52 different HCV isolates. Isolated cDNA sequences are used for the prodn. of primers useful for detecting the presence of HCV in a sample, the primers are also useful for HCV genotyping. Proteins encoded by the cdnas can be used

Db 1 MSTITPKPQKTKRNTNRPRDVKPGGQIVGGVYIIPRGRGLGVTRKTSRSQPRG 60
 Qy 61 RQPQPKQPKQPOQGRHWAQPGVPPWLYKGNGCAGWMLSPRSRPRPNSRLG 120
 Db 61 RQPQPKQPKQPOQGRHWAQPGVPPWLYKGNGCAGWMLSPRSRPRPNSRLG 120
 Qy 121 KVIDLTTCGFADLGYTPVVGAPLGGAALAHGVRAEDGINVATGNIPGCSSFSIPLLA 180
 Db 121 KVIDLTTCGFADLGYTPVVGAPLGGAALAHGVRAEDGINVATGNIPGCSSFSIPLLA 180
 Qy 181 LISCLTPASA 191
 Db 181 LISCLTPASA 191

RESULT 10
 ID AAR7447 standard; protein; 196 AA.
 XX
 XX AAR74047;
 DT 25-MAR-2003 (revised)
 DT 26-NOV-1995 (first entry)
 XX Synthetic HCV nucleocapsid protein.
 XX Hepatitis C virus; vector; expression; prokaryotic cells; detection;
 KW diagnosis.
 XX Synthetic.
 DE Synthetic HCV nucleocapsid protein.
 XX
 KW Hepatitis C virus; subtype; polymerase chain reaction; amplification;
 KW PCR; primer; probe; antibody; infection.
 XX
 OS Hepatitis C virus.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 128
 FT /label= Phe, Ser, Tyr, Cys
 FT
 FT Misc-difference 129
 FT /label= Gly
 FT /note= "amino acid in this position is designated X in the specific sequence, but codon usage shows that the only possible amino acid at this pos. is Gly"
 FT Misc-difference 308
 FT /label= Ile, Met
 FT
 FT Misc-difference 310
 FT /label= Ser, Pro, Thr, Ala
 XX
 PR WO9511980-A2.
 XX
 PR 25-OCT-1993; 93US-00141917.
 XX
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 XX
 PT Khudyakov Y, Fields HA;
 DR WPI; 1995-178872/23.
 XX
 PT Vector for expression of a synthetically produced protein coding DNA -
 PR pref. encodes the hepatitis C virus nucleo-capsid protein which can be
 XX used in the detection of HCV antibodies.
 PS Disclosure; Page 35; 44pp; English.

XX The sequence is that of a synthetic nucleotide sequence encoding the
 CC hepatitis C virus nucleocapsid protein. The gene is positioned in a
 CC vector for efficient expression in prokaryotic cells. The expressed
 CC protein can be used in tests for the detection of antibodies specific for
 CC the HCV C protein. See also AAR74048. (Updated on 25-MAR-2003 to correct
 CC
 XX
 SQ Sequence 196 AA;

Query Match 93.5%; Score 977; DB 2; Length 196;
 Best Local Similarity 92.1%; Pred. No. 6.1e-84;
 Matches 176; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

Qy 1 MSTITPKPQKTKRNTNRPRDVKPGGQIVGGVYIIPRGRGLGVTRKTSRSQPRG 60
 1 MSTITPKPQKTKRNTNRPRDVKPGGQIVGGVYIIPRGRGLGVTRKTSRSQPRG 60
 Qy 61 RQPQPKQPKQPOQGRHWAQPGVPPWLYKGNGCAGWMLSPRSRPRPNSRLG 120
 61 RQPQPKQPKQPOQGRHWAQPGVPPWLYKGNGCAGWMLSPRSRPRPNSRLG 120
 Db 61 RQPQPKQPKQPOQGRHWAQPGVPPWLYKGNGCAGWMLSPRSRPRPNSRLG 120

RESULT 11
 ID AAR96546 standard; peptide; 319 AA.
 XX
 XX AAR96546;
 DT 10-MAR-1997 (first entry)
 XX
 DE Hepatitis C virus types 9a(7a) isolates FR1 amino acids 1-317.
 XX
 KW Hepatitis C virus; subtype; polymerase chain reaction; amplification;
 KW PCR; primer; probe; antibody; infection.
 XX
 OS Hepatitis C virus.

XX
 FH Key Location/Qualifiers
 FT Misc-difference 128
 FT /label= Phe, Ser, Tyr, Cys
 FT
 FT Misc-difference 129
 FT /label= Gly
 FT /note= "amino acid in this position is designated X in the specific sequence, but codon usage shows that the only possible amino acid at this pos. is Gly"
 FT Misc-difference 308
 FT /label= Ile, Met
 FT
 FT Misc-difference 310
 FT /label= Ser, Pro, Thr, Ala
 XX
 PR WO9613590-A2.
 XX
 PR 09-MAY-1996.
 XX
 PR 23-OCT-1995; 95WO-EP001155.
 XX
 PR 21-OCT-1994; 94EP-00870166.
 PR 28-JUN-1995; 95EP-00870076.
 XX
 PA (INNO-) INNOGENETICS NV.
 XX
 PT Maertens G, Stuyver L;
 XX
 DR WPI; 1996-251460/25.
 XX
 N-PSDB; AAN27957.

XX
 PR Hepatitis C virus poly:nucleic acid unique to unidentified sub:type -
 PR used to develop probes and primers for new sub:types and vaccines to
 PR prevent and treat infection.
 XX
 PS Claim 25; Fig 3; 150pp; English.

XX
 CC The sequences AAR96546-R96578 represent novel sequences isolated from
 CC hepatitis C virus subtypes different from subtypes 1a-c, 2a-d, 3a-f, 4a-
 CC j, 5a and 6a. They esp. from the novel subtypes 1d-f, 2e-i, 2k, 2l, 3g,
 CC 4k-m, 7a-c or types 9, 10 or 11. The sequences correspond to the 5'-
 CC untranslated region (UR), Core/EL1, NS4 or NS5 regions of the genome.
 CC This sequence represents amino acids 1-317 from the HCV types 9a and 7a
 isolates FRI. The new HCV types were isolated from patients with chronic
 CC HCV from the Benelux countries, France, Cameroon and Vietnam, because of
 CC their aberrant reactivities. The RNA was extracted, cDNA synthesised and
 CC PCR amplified, cloned and genotyped. The 5'UR, Core/EL1 and NS5 regions
 CC were sequenced either directly or partially and used to classify the new
 CC viruses into (sub)types based on comparison with known sequences. The
 CC sequences were used to generate the peptides AAR96424-R96524. The

CC sequences can also be used to synthesise probes and primers for the detection of HCV in a sample. The polypeptides can be used to detect anti HCV antibodies, for HCV typing or to prevent HCV infections

CC

CC HCV antibodies, for HCV typing or to prevent HCV infections

XX

SQ Sequence 319 AA;

Query Match Best Local Similarity 93.5%; Score 977; DB 2; Length 319; Matches 179; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

Ov 1 MSTLPKPKRKTKRNTNRPPTDKPVGCGQIVGGVYLPRRGPRIGRATKTSERSOPRG 60
Db 1 MSTLPKPKRKTKRNTNRPPTDKPVGCGQIVGGVYLPRRGPRIGRATKTSERSOPRG 60

Ov 61 RROPIPKARQPOGRHWAQPGIPWPLYNGEGGWAGMILSPRSRSPWGPNDRRSRLG 120
Db 61 RROPIPKARQPOGRHWAQPGIPWPLYNGEGGWAGMILSPRSRSPWGPNDRRSRLG 120

Ov 121 KVIDTLLCFLADLMLGYIPVLGVGPIGGVAAALAHGVRAIEDGVNVATGNLPGCSFSIFLA 180
Db 121 KVIDTLLCFLADLMLGYIPVLGVGPIGGVAAALAHGVRAIEDGVNVATGNLPGCSFSIFLA 180

Ov 181 LLSCLTPASA 191
Db 181 LLSCLTPASA 191

RESULT 12

ID AAR22137 standard; protein: 326 AA.

AC AAR22137;

XX

DT 27-AUG-2003 (revised)
DT 25-MAR-2003 (revised)
DT 07-JUL-1992 (first entry)

XX

DE ICSV-Hc59 capsid and envelope proteins.

XX

KW Hepatitis C virus; non-A non-B virus; HCV-Hc59; antigen; vaccine; assay; detection.

XX

OS Non-A, non-B hepatitis virus.

FH Region

Key Location/Qualifiers

FT .326 NANBV_structural_proteins

FT .326 /label= NANBV_structural_proteins

FT Peptide 1..120 /label= capsid

FT Peptide 1..74 /label= pref._capsid_antigen

FT Peptide 1..20 /label= pref._capsid_antigen

FT Peptide 2..40 /label= pref._capsid_antigen

FT Peptide 3 /label= pref._capsid_antigen

FT Misc-difference 3 /label= "or Asn according to Seq No 46 (AAR22154), see CC"

FT Peptide 21..40 /label= pref._capsid_antigen

FT Peptide 69..120 /label= pref._capsid_antigen

FT Region 121..326 /label= envelope_protein

FT Peptide 121..176 /label= pref._envelope_antigen

FT Misc-difference 321 /label= Asp /note= "or Asn according to Seq No 1 (AAR22154), see CC"

XX W09203458-A.

XX PN

PD 05-MAR-1992.

XX ID AAR38278

XX DT 23-AUG-1991; 91W0-US006037.

XX PR 25-AUG-1990; 90US-00573643.

XX PR 21-NOV-1990; 90US-00616369.

XX PR 21-AUG-1991; 91US-00748564.

PA (NYBL) NEW YORK BLOO DCENT.

PA (PHAR-) PHARMA.

PT Zebedee S, Inchauspe G, Nasofo MS, Prince AM;

PT XX WPI; 1992-096821/12.

DR DR N-PSDB; AAQ22238.

PT Deoxyribonucleic acid sequence encoding polypeptide(s), useful as vaccines, and obtained from C59 subgroup encoding polypeptide(s), useful as vaccines, and immuno reactive ABS for diagnosis of virus.

PS Disclosure; Page 131-133; 225pp; English.

XX One Hutch strain (HCV-H) of NANBV, designated the Hutch C59 isolate (HCV-Hc59) was propagated through passage in animals and the entire viral genome was cloned and sequenced (see AAR22871). The sequence represented here comprises two amino acids, indicated in the features, which differ from the sequence of AAR22154. The proteins and peptides (see features) and antibodies against them are useful for the prep. of vaccines and inoculants against NANBV and in immunological assays for detection of viral infection. (Updated on 25-MAR-2003 to correct PR field.) (Updated on 25-MAR-2003 to correct CC field.) (Updated on 27-AUG-2003 to correct CC field.) (Updated on 27-AUG-2003 to correct OS field.)

SQ Sequence 326 AA;

Query Match Best Local Similarity 92.1%; Score 977; DB 2; Length 326; Matches 176; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

Ov 1 MSTLPKPKRKTKRNTNRPPTDKPVGCGQIVGGVYLPRRGPRIGRATKTSERSOPRG 60
Db 1 MSTLPKPKRKTKRNTNRPPTDKPVGCGQIVGGVYLPRRGPRIGRATKTSERSOPRG 60

Ov 61 RROPIPKARQPOGRHWAQPGIPWPLYNGEGGWAGMILSPRSRSPWGPNDRRSRLG 120
Db 61 RROPIPKARQPOGRHWAQPGIPWPLYNGEGGWAGMILSPRSRSPWGPNDRRSRLG 120

Ov 121 KVIDTLLCFLADLMLGYIPVLGVGPIGGVAAALAHGVRAIEDGVNVATGNLPGCSFSIFLA 180
Db 121 KVIDTLLCFLADLMLGYIPVLGVGPIGGVAAALAHGVRAIEDGVNVATGNLPGCSFSIFLA 180

Ov 181 LLSCLTPASA 191
Db 181 LLSCLTPASA 191

RESULT 13

ID AAR38278

XX ID AAR38278 standard; protein: 733 AA.

XX AC AAR38278;

XX DT 21-OCT-1993 (first entry)

XX DE NANB hepatitis virus HC-OM gene polypeptide P-733-1.

XX KW Non-A, non-B; virus; polymerase chain reaction; detection; sensitive;

XX KW specific; HCV; NANB.

OS Non-A hepatitis virus.

XX PN JP05091884-A.

XX
XX PD 16-APR-1993.
XX FT
XX PR 10-APR-1991; 91JP-00196175.
XX FT
XX PR 12-JUN-1990; 90JP-00153401.
XX PR 08-NOV-1990; 90JP-00304405.
XX PA (NAKA,/) NAKAMURA T.
XX DR WPI; 1993-199637/25.
XX DR N-PSDB; A4Q43888.
XX PS Claim 14; Page 32-35; 73pp; Japanese.

XX PT Antigen related to non-A and non-B hepatitis virus - comprises non-
PT translation region comprising 340 - 341 mols. of nucleotides non-
PT translation region comprising 1885 - 2551 mols. of nucleotides including
PT region 1,149 and, etc.
XX CC
XX SQ Sequence 733 AA;

Query Match 93.5%; Score 977; DB 2; Length 733;
Best Local Similarity 92.1%; Pred. No. 2.7e-83; Pairs 6;
Matches 176; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

QY 1 MSTLKPORKTKRNTNRPTDKFPGGGQIVGCVYLPRRGRLGVTRKTSERSOPRG 60
Db 1 MSTIPKPKRKTKRNTNRPTDKFPGGGQIVGCVYLPRRGRLGVTRKTSERSOPRG 60

QY 61 RQQPPIPKARQDRCRHAWQPGYVPLYNGNEBGCCWAGWLLSPRSRPRHGPNDRRSRNLG 120
Db 61 RQQPPIPKVRRPSEGRTWAAQPGYVPLYNGNEBGCCWAGWLLSPRSRPSNGPTDRRRSRNLG 120

QY 121 KVIDTLLCGFAIDMGYIPLVGAATLGGAARALANGVRVIEDGNVATGNLPGSFSITLA 180
Db 121 KVIDTLLCGFAIDMGYIPLVGAATLGGAARALANGVRVIEDGNVATGNLPGSFSITLA 180

QY 181 LISCLTPASA 191
Db 181 LISCLTPASA 191

FT Peptide /label= 5
FT Peptide /label= 6
FT Peptide /label= 7
FT Peptide /label= 7
FT Peptide /label= 8
FT Peptide /label= 9
FT Peptide /label= 9
FT Peptide /label= 10
FT Peptide /label= 10
FT Peptide /label= 11
FT Peptide /label= 11
FT Peptide /label= 11
FT Peptide /label= 12
FT Peptide /label= 12
FT Peptide /label= 13
FT Peptide /label= 13
FT Peptide /label= 14
FT Peptide /label= 14
FT Peptide /label= 14
FT Peptide /label= 15
FT Peptide /label= 15
FT Peptide /label= 16
FT Peptide /label= 16
FT Peptide /label= 17
FT Peptide /label= 17
FT Peptide /label= 18
FT Peptide /label= 18
FT Peptide /label= 19
FT Peptide /label= 19
PN EP489568-A1.

XX PD 17-JUN-1992.

XX PR 14-DEC-1990; 90EP-00124241.

XX PR (INNO-) INNOGENETICS NV.

XX PI Deleys RJ, Pollet D, Maertens G, Van Heuverswyn H;

XX DR WPI; 1992-201383/25.

XX PR New synthetic peptides for detecting antibodies to hepatitis C virus -

PT useful in e.g. ELISA assays, and for detection of HCV antigens or as

PT immunogens.

XX PS Disclosure; Fig 1; 32pp; English.

XX CC RNA viruses frequently exhibit a high rate of spontaneous mutation, thus

CC a virus is considered to be the same of equiv. to HCV if it exhibits a
CC global homology of more than 70 percent with the HCV HC-J1/CDC/CHI

CC composite sequence. The peptide fragments of this DNA sequence indicated

CC in the features table can immunologically mimic proteins encoded by HCV.

CC Additional amino acids or chemical gags, may be added to either end of the

CC peptides for the purpose of creating a linker arm for attachment to a

CC carrier. The peptides may be used for the detection of antibodies

CC specific for HCV. They may be used in the form of kits, opt. with

CC reagents such as staphylococcal protein A, streptococcal protein G,

CC avidin or streptavidin. The peptides may also be used as immunogens for

CC raising antibodies. (Updated on 25-MAR-2003 to correct PN field.)

XX SQ Sequence 2894 AA;

XX PS Query Match 93.5%; Score 977; DB 2; Length 2894;
Best Local Similarity 92.1%; Pred. No. 1.2e-82; Pairs 6;
Matches 176; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

QY 1 MSTLKPORKTKRNTNRPTDKFPGGGQIVGCVYLPRRGRLGVTRKTSERSOPRG 60
Db 1 MSTIPKPKRKTKRNTNRPTDKFPGGGQIVGCVYLPRRGRLGVTRKTSERSOPRG 60

| | | | |
|----|---|----|---|
| AC | AAR44010; | QY | 181 LLSCLTPASA 191 |
| XX | | Db | 181 LLSCLTPASA 191 |
| DT | 25-MAR-2003 (revised) | ID | AAR2938 Standard; protein; 191 AA. |
| DT | 12-MAY-1994 (first entry) | XX | |
| DE | Hepatitis C virus core protein. | AC | AAR2938; |
| XX | | XX | |
| KW | HCV; non-A, non-B hepatitis virus; NANBH; immunodominant region; capsid protein; vaccine; antigen. | XX | |
| XX | | XX | |
| OS | Hepatitis C virus. | OS | |
| XX | | XX | |
| FT | Location/Qualifiers | XX | |
| FT | Peptide | DE | Hepatitis C virus isolate S14 core protein. |
| FT | 1. .45 /label= S42G /note= "peptide 1" | XX | |
| FT | 1. .45 /label= immunodominant region /note= "novel antigenic peptides contain sequences from this region" | KW | HCV; El; envelope 1; core protein; HCV genotyping; antibody; vaccine; hepatitis. |
| FT | Peptide | XX | |
| FT | Peptide | XX | |
| FT | 1. .45 /label= P42Y /note= "peptide 2" | XX | |
| FT | 77. .115 /label= R40R /note= "peptide 3" | XX | |
| PP | 06-MAY-1993; 93EP-00420183. | PP | 15-AUG-1995; 95WO-US010398. |
| PP | 06-MAY-1992; 92FR-00005763. | PP | 15-AUG-1994; 94US-00290665. |
| PR | | PD | 22-FEB-1996. |
| XX | | XX | |
| PN | EP569309-A1. | PA | (USSH) US SEC DEPT HEALTH. |
| XX | | XX | |
| PD | 10-NOV-1993. | PI | Bukh J, Miller RH, Purcell RH; |
| XX | | XX | |
| PT | Dalbon P, Jolivet M; | PS | WPI; 1996-139709/14. |
| XX | | DR | N-PSDB; AATI6612. |
| DR | | XX | |
| XX | | XX | |
| PA | (INMR) BIO MERIRUX. | PT | DNA and amino acid sequence of HCV envelope 1 and core proteins - used to determine HCV genotype and as vaccines against HCV infection. |
| XX | | XX | |
| PT | New antigenic polypeptide(s) from hepatitis C virus - derived from N-terminal region of core protein, and related antibodies, useful in diagnosis, vaccination and treatment. | PS | Claim 4; Page 182; 340pp; English. |
| XX | | XX | |
| PT | New antigenic polypeptide(s) from hepatitis C virus - derived from N-terminal region of core protein, and related antibodies, useful in diagnosis, vaccination and treatment. | CC | AAR2936-R22987 are HCV core proteins derived from 52 different HCV isolates. Isolated cDNA sequences are used for the prodn. of primers useful for detecting the presence of HCV in a sample, the primers are also useful for HCV genotyping. Proteins encoded by the cDNAs can be used in vaccines for immunising against HCV infection. The proteins may also be used to detect antibodies against HCV in serum, saliva, lymphocytes or other mononuclear cells. The antibodies may be used in the prevention of HCV infection. |
| XX | | CC | |
| CC | immunodominant region of HCV core protein. Further analysis of the N-terminal 45 amino acids of the HCV core protein showed that some epitopes are present in the first 21 amino acids, some in 22-45 and some at the junction of these two subregions. Preferred antigenic peptides are derived from S42G; they are useful for detecting HCV antibodies and in vaccines against HCV. (Updated on 25-MAR-2003 to correct PN field.) | CC | |
| CC | | CC | |
| CC | of peptides 1-3 (see Features Table), peptide 1 was found to be from the | CC | |
| CC | terminal 45 amino acids of the HCV core protein. Further analysis of the N-terminal 45 amino acids of the HCV core protein showed that some epitopes are present in the first 21 amino acids, some in 22-45 and some at the junction of these two subregions. Preferred antigenic peptides are derived from S42G; they are useful for detecting HCV antibodies and in vaccines against HCV. (Updated on 25-MAR-2003 to correct PN field.) | CC | |
| XX | Sequence 191 AA; | XX | |
| SQ | Query Match 93.4%; Score 976; DB 2; Length 191; Best Local Similarity 92.7%; Pred. No. 7. 4e-84; Mismatches 9; Indels 0; Gaps 0; Matches 177; Conservative | QY | 93.4%; Score 976; DB 2; Length 191; Best Local Similarity 92.7%; Pred. No. 7. 4e-84; Mismatches 9; Indels 0; Gaps 0; Matches 177; Conservative |
| QY | 1 MSTLPKPKRTKNTNRPTDKFPGGGQIVGGYLLPRLRRGRPLGRTRAKTSERSOPRG 60 | QY | 1 MSTLPKPKRTKNTNRPTDKFPGGGQIVGGYLLPRLRRGRPLGRTRAKTSERSOPRG 60 |
| Db | 1 MSTNPKPQPKTRKNTNRPTDKFPGGGQIVGGYLLPRLRRGRPLGRTRAKTSERSOPRG 60 | Db | 61 RROPIPKARRQPOGRHWAQPGYWPMLYGNBGCWAGWLLSPRSRPHKGPNPDRRSRNLG 120 |
| QY | 61 RROPIPKARRQPOGRHWAQPGYWPMLYGNBGCWAGWLLSPRSRPHKGPNPDRRSRNLG 120 | QY | 61 RROPIPKARRQPOGRHWAQPGYWPMLYGNBGCWAGWLLSPRSRPHKGPNPDRRSRNLG 120 |
| Db | 121 KVIDLTGCFADIMGYIPVGAPLGGAVALAHLGVRAIEDGINYATCNLPGCSFSFLIA 180 | Db | 121 KVIDLTGCFADIMGYIPVGAPLGGAVALAHLGVRAIEDGINYATCNLPGCSFSFLIA 180 |
| QY | 121 KVIDLTGCFADIMGYIPVGAPLGGAVALAHLGVRAIEDGINYATCNLPGCSFSFLIA 180 | QY | 181 LISCLTPASA 191 |
| Db | 121 KVIDLTGCFADIMGYIPVGAPLGGAVALAHLGVRAIEDGINYATCNLPGCSFSFLIA 180 | Db | 181 LISCLTPASA 191 |

XX AAR92941 Hepatitis C virus isolate US11 core protein.
 ID AAR92941 Standard; protein; 191 AA.
 XX
 AC AAR92941;
 XX
 DT 02-OCT-1996 (first entry)
 XX
 DE Hepatitis C virus isolate DRA core protein.
 XX
 KW HCV; E1; envelope 1; core protein; HCV genotyping; antibody; vaccine;
 hepatitis.
 XX
 OS Hepatitis C virus.
 XX
 PN WO9605315-A2.
 XX
 XX 22-FEB-1996.
 PD
 XX
 PR 15-AUG-1995; 95WO-US010398.
 XX
 PR 15-AUG-1994; 94US-00290665.
 XX
 PA (USSH) US SEC DEPT HEALTH.
 XX
 PT Bukh J, Miller RH, Purcell RH;
 XX
 DR WPI; 1996-139709/14.
 XX
 DR N-PSDB; AAT16615.
 XX
 PT DNA and amino acid sequence of HCV envelope 1 and core proteins - used to
 determine HCV genotype and as vaccines against HCV infection.
 XX
 PR DNA and amino acid sequence of HCV envelope 1 and core proteins - used to
 determine HCV genotype and as vaccines against HCV infection.
 XX
 PS Claim 4; Page 184-185; 340pp; English.
 XX
 CC AAR92936-R92987 are HCV core proteins derived from 52 different HCV
 isolates. Isolated cDNA sequences are used for the prodn. of primers
 useful for detecting the presence of HCV in a sample, the primers are
 also useful for HCV genotyping. Proteins encoded by the cDNAs can be used
 in vaccines for immunising against HCV infection. The proteins may also
 be used to detect antibodies against HCV in serum, saliva, lymphocytes or
 other mononuclear cells. The antibodies may be used in the prevention of
 HCV infection.
 XX
 SQ Sequence 191 AA;
 Query Match 93.4%; Score 976; DB 2; Length 191;
 Best Local Similarity 92.7%; Pred. No. 7.4e-84;
 Mismatches 5; Indels 0; Gaps 0;
 Matches 177; Conservative 9;
 QY 1 MSTLPKPKPDKTRNTRPRDVKFPGGQIVGGVILPRRGRPLGRATRKSERSQPRG 60
 Db 1 MSTLPKPKPDKTRNTRPRDVKFPGGQIVGGVILPRRGRPLGRATRKSERSQPRG 60
 QY 61 RQQPKRQPKRQPRHWAQPGYPWMLYQNGCGWAGWLSPRGSPBPHWGNDPRRSRNLG 120
 Db 61 RQQPKRQPKRQPRHWAQPGYPWMLYQNGCGWAGWLSPRGSPBPHWGNDPRRSRNLG 120
 QY 121 KVIDLTGCGADLGMGYIPVGAPLGVAALAHGRAEDGINVATGNLPCCSFSIFLLA 180
 Db 121 KVIDLTGCGADLGMGYIPVGAPLGVAALAHGRAEDGINVATGNLPCCSFSIFLLA 180
 QY 181 LLSCLTPASA 191
 Db 181 LLSCLTPASA 191
 QY 181 LLSCLTPASA 191
 Db 181 LLSCLTPASA 191
 RESULT 21
 AAR92939 Hepatitis C virus isolate SW1 core protein.
 ID AAR92939 Standard; protein; 191 AA.
 XX
 AC AAR92939;
 XX
 DT 02-OCT-1996 (first entry)
 XX
 DE Hepatitis C virus isolate SW1 core protein.
 XX
 KW HCV; E1; envelope 1; core protein; HCV genotyping; antibody; vaccine;
 hepatitis.
 XX

OS Hepatitis C virus.
 XX WO9605315-A2.
 XX PD 22-FEB-1996.
 XX PP 15-AUG-1995; 95WO-US010398.
 XX PR 15-AUG-1994; 94US-00230665.
 XX PA (USSH) US SEC DEPT HEALTH.
 XX PI Bukh J, Miller RH, Purcell RH;
 XX DR WPI: 1996-139709/14.
 XX N-PSDB; AAT1613.
 XX PT DNA and amino acid sequence of HCV envelope 1 and core proteins - used to determine HCV genotype and as vaccines against HCV infection.
 XX Claim 4; Page 183; 340pp; English.
 CC AR92935-R2987 are HCV core proteins derived from 52 different HCV isolates. Isolated cDNA sequences are used for the prodn. of primers useful for detecting the presence of HCV in a sample, the primers are also useful for HCV genotyping. Proteins encoded by the cDNAs can be used in vaccines for immunising against HCV infection. The proteins may also be used to detect antibodies against HCV in serum, saliva, lymphocytes or other mononuclear cells. The antibodies may be used in the prevention of HCV infection.
 XX Sequence 191 AA;
 SQ Query Match 93.4%; Score 976; DB 2; Length 191;
 Best Local Similarity 92.7%; Pred. No. 7.4e-84;
 Matches 177; Conservative 5; Mismatches 9; Indels 0; Gaps 0;
 CC 1 MSTLPKPKRTKENTNRPTDKPFGQQIVGGYLLPRRGRLGVATRKTSERSOPRG 60
 Db 1 MSTNPQPKQTKTNTNRPKDQKTFPGGQIVGGYLLPRGRGLGVATRKTSERSOPRG 60
 QY 61 RQQPPIKARQPOCRHWAQPGYIPLWYNEGGCWAGWILLSPRSRPRHGPNDRRSRNLG 120
 Db 61 RQQPPIKARRPEGRTWAPGYPWPLWYNEGGCWAGWILLSPRSRSPWGPNDPRRSRN LG 120
 QY 121 KVIDTLTGFAIDLGYIPVUGAHLGGVAALAHGVRAIEDGNYATENLPGCSFSILLA 180
 Db 121 KVIDTLTGFAIDLGYIPVUGAHLGGVAALAHGVRAIEDGNYATENLPGCSFSILLA 180
 QY 181 LISCLTPASA 191
 Db 181 LISCLTVASA 191
 RESULT 22
 AR9240 AAY94411 standard; protein; 191 AA.
 ID AR92940 standard; protein; 191 AA.
 XX AC AAR92940;
 XX DT 02-OCT-1996 (first entry)
 XX Hepatitis C virus isolate S18 core protein.
 DE Hepatitis C virus envelope 1; core protein; HCV genotyping; antibody; vaccine;
 KW hepatitis.
 OS Hepatitis C virus.
 XX HCV; El: envelope 1; core protein; HCV genotyping; antibody; vaccine;
 KW hepatitis.
 OS Hepatitis C virus.
 XX PN WO9605315-A2.
 XX PD 22-FEB-1996.
 XX
 PF 15-AUG-1995; 95WO-US010398.
 XX PR 15-AUG-1994; 94US-00230665.
 XX PA (USSH) US SEC DEPT HEALTH.
 XX PI Bukh J, Miller RH, Purcell RH;
 XX DR WPI: 1996-139709/14.
 XX N-PSDB; AAT1614.
 XX PT DNA and amino acid sequence of HCV envelope 1 and core proteins - used to determine HCV genotype and as vaccines against HCV infection.
 XX Claim 4; Page 184; 340pp; English.
 CC AR92935-R2987 are HCV core proteins derived from 52 different HCV isolates. Isolated cDNA sequences are used for the prodn. of primers useful for detecting the presence of HCV in a sample, the primers are also useful for HCV genotyping. Proteins encoded by the cDNAs can be used in vaccines for immunising against HCV infection. The proteins may also be used to detect antibodies against HCV in serum, saliva, lymphocytes or other mononuclear cells. The antibodies may be used in the prevention of HCV infection.
 XX Sequence 191 AA;
 SQ Query Match 93.4%; Score 976; DB 2; length 191;
 Best Local Similarity 92.7%; Pred. No. 7.4e-84;
 Matches 177; Conservative 5; Mismatches 9; Indels 0; Gaps 0;
 CC 1 MSTLPKPKRTKENTNRPTDKPFGQQIVGGYLLPRRGRLGVATRKTSERSOPRG 60
 Db 1 MSTNPQPKQTKTNTNRPKDQKTFPGGQIVGGYLLPRGRGLGVATRKTSERSOPRG 60
 QY 61 RQQPPIKARQPOCRHWAQPGYIPLWYNEGGCWAGWILLSPRSRPRHGPNDRRSRNLG 120
 Db 61 RQQPPIKARRPEGRTWAPGYPWPLWYNEGGCWAGWILLSPRSRSPWGPNDPRRSRN LG 120
 QY 121 KVIDTLTGFAIDLGYIPVUGAHLGGVAALAHGVRAIEDGNYATENLPGCSFSILLA 180
 Db 121 KVIDTLTGFAIDLGYIPVUGAHLGGVAALAHGVRAIEDGNYATENLPGCSFSILLA 180
 QY 181 LISCLTPASA 191
 Db 181 LISCLTVASA 191
 RESULT 23
 AR9240 AAY94411 standard; peptide; 191 AA.
 ID AAY94411 standard; peptide; 191 AA.
 XX AC AAY94411;
 XX DT 11-SEP-2000 (first entry)
 XX DE Human hepatitis C virus core protein.
 XX KW Human; hepatitis C virus; HCV; immunogen; antigen; antibody; virucide; hepatotropic; anti-inflammatory; virus detection; vaccine.
 XX OS Hepatitis C virus.
 XX Hepatitis C virus.
 PN WO200031130-A1.
 XX PD 02-JUN-2000.
 XX PR 19-NOV-1999; 99WO-1B001933.
 XX PR 20-NOV-1998; 98US-00196155.
 XX PA (INMR) BIO MERIEUX.

PI Dalbon P, Jolivet M, Jolivet-Reynaud C;
 XX DR WPI; 2000-411934/35.
 XX PT Polypeptides that bind to anti-hepatitis C virus antibodies, useful for
 PT diagnosing and preventing hepatitis C infections.
 XX PS Disclosure; Fig 1; 50pp; English.

CC The present sequence is the core protein of the human hepatitis C virus (HCV). Three long polypeptide fragments of about 40 amino acids in the 120 N-terminal amino acids of this sequence were synthesised and their reactivity towards HCV-positive sera was evaluated in an ELISA test. The reactivity of the first peptide, S42G, which extends from serine 2 up to glycine 45, was substantially greater than the reactivity of the other peptides. Peptide S42G represents an immunodominant region containing conformational type epitopes and linear-type epitopes. S42G manifests an immunoreactivity with all the sera of individuals or blood samples infected with HCV and which possess antibodies directed against the core protein. An amino acid of the S42G sequence may be substituted for homologous amino acids and side chains and peptide bonds may also be modified. For example, L-amino acids may be replaced by D-amino acids, amine groups may be acetylated, and so on. The native antigenic sequence and its antigenic derivatives may be used for detection of human hepatitis C viruses and for raising antibodies against the virus

XX SQ Sequence 191 AA;

Query Match Best Local Similarity 93.4%; Score 976; DB 3; Length 191; Matches 177; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

Qy 1 MSTLPKPKRTKENTNRPRPTDKPFGGGQIVGGVYLPRRGRLGRATRKTSERSOPRG 60
 Db 1 MSTNPKPQPKRTKNTNRPRQDKPQTKPFGGGQIVGGVYLPRRGRLGRATRKTSERSOPRG 60

Qy 61 RQDPIPKARQPOCRHWAQPGYWPWLYPNEGCWAGWGLSPRSRPHGNDPDRRSNLG 120
 Db 61 RQDPIPKARQPOCRHWAQPGYWPWLYPNEGCWAGWGLSPRSRPHGNDPDRRSNLG 120

Qy 121 KVIDTLLTCGFAIDMGYIPVVGAPLGVAALAHGVATEDGINYATGNLPGCSFSIPLA 180
 Db 121 KVIDTLLTCGFAIDMGYIPVVGAPLGVAALAHGVATEDGINYATGNLPGCSFSIPLA 180

Qy 181 LSLCLTVPASA 191
 Db 181 LSLCLTVPASA 191

RESULT 24

BAR67591 ID AR67591 standard; protein; 502 AA.
 XX AC AAR67591;
 XX DT 07-SEP-1995 (first entry)
 XX DE Hepatitis C virus SR037-5' gene product.
 XX KW Hepatitis C virus; HCV; non-A non-B; SR037-5'; treatment.
 XX OS Hepatitis C virus.
 XX PN JP06319563-A.
 XX PD 22-NOV-1994.
 XX PP 13-MAY-1993; 93JP-00147133.
 XX PR 13-MAY-1993; 93JP-00147133.
 XX PA (IMMO) IMMUNO JAPAN KK.

DR WPI; 1995-040318/06.
 DR N-PSDB; AAQ79143.
 XX PT A hepatitis C virus gene and oligo-nucleotide(s) - used for the treatment of hepatitis C.
 XX PT Claim 21; Page 36; 41pp; Japanese.

CC AAQ79143 is the hepatitis C virus (HCV) gene SR037-5', cDNA, it encodes the protein described in AR67591. Both the cDNA and protein can be used in the treatment of HCV infection

XX SQ Sequence 502 AA;

Query Match Best Local Similarity 92.1%; Score 976; DB 2; Length 502; Matches 176; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

Qy 1 MSTLPKPKRTKENTNRPRPTDKPFGGGQIVGGVYLPRRGRLGRATRKTSERSOPRG 60
 Db 1 MSTNPKPQPKRTKNTNRPRQDKPQTKPFGGGQIVGGVYLPRRGRLGRATRKTSERSOPRG 60

Qy 61 RQDPIPKARQPOCRHWAQPGYWPWLYPNEGCWAGWGLSPRSRPHGNDPDRRSNLG 120
 Db 61 RQDPIPKARQPOCRHWAQPGYWPWLYPNEGCWAGWGLSPRSRPHGNDPDRRSNLG 120

Qy 121 KVIDTLLTCGFAIDMGYIPVVGAPLGVAALAHGVATEDGINYATGNLPGCSFSIPLA 180
 Db 121 KVIDTLLTCGFAIDMGYIPVVGAPLGVAALAHGVATEDGINYATGNLPGCSFSIPLA 180

Qy 181 LSLCLTVPASA 191
 Db 181 LSLCLTVPASA 191

RESULT 25

AR67922 ID AR67922 standard; protein; 967 AA.
 XX AC AAR7922;
 XX DT 08-DEC-1995 (first entry)
 XX DB DHCV11-encoded sequence.
 XX KN PHCV11; amyloid precursor protein; APP; hepatitis C virus; HCV; E1; E2; fusion protein; HEK-293; vaccine; vector; prc/CMV; protein secretion; glycosylation.
 XX OS Synthetic.
 XX PN W09520664-A2.
 XX PD 03-AUG-1995.
 XX PF 27-JAN-1995; 95WO-US001087.
 XX PR 28-JAN-1994; 94US-00188281.
 XX PA (ABBO) ABBOTT LAB.
 XX PI Watanabe S, Yamaguchi J, Desai SM, Devare SG;
 XX DR WPI; 1995-275449/36.
 XX PT New mammalian expression systems for HCV proteins - express fusion protein comprising amyloid precursor protein and HCV B1 and/or B2 protein.
 XX Disclosure; Page 59-62; 89pp; English.

CC RNA from serum or plasma of a chimpanzee infected with HCV was converted to cDNA and PCR amplified using primers based on HCV sequences. 7

CC Adjacent HCV DNA fragments were generated which together encoded the HCV
 CC sequence given in AAR79232. Fragments from 2 clones, pICV141 and pICV150
 CC (see AAR79223), were combined to generate PHCV176 (see AAR79221). This
 CC was used to construct APP-HCV E1+E2 fusion proteins that were
 CC glycosylated and secreted from HEK-293 transfecant^g

XX SQ Sequence 967 AA;

Query Match 93.4%; Score 976; DB 2; Length 967;
 Best Local Similarity 92.7%; Pred. No. 4 5e-83; 5; Mismatches 9; Indels 0; Gaps 0;
 Matches 177; Conservative

Qy 1 MSTLPKPKRKTKRNTNRPTDVKPFPGQQIVGGVYLPRRGPRGRGTRAKTSRSOPRG 60
 1 MSTLPKPKRKTKRNTNRPTDVKPFPGQQIVGGVYLPRRGPRGRGTRAKTSRSOPRG 60
 Db 61 RROPIPKARRQPOGRHWAQPGYPPWPLYGNBEGCGWAGWILSPRSRPHGPNDRRSRNLG 120
 Qy 61 RROPIPKARRQPOGRHWAQPGYPPWPLYGNBEGCGWAGWILSPRSRPHGPNDRRSRNLG 120
 61 RROPIPKARRQPOGRHWAQPGYPPWPLYGNBEGCGWAGWILSPRSRPHGPNDRRSRNLG 120
 Db 61 RROPIPKARRQPOGRHWAQPGYPPWPLYGNBEGCGWAGWILSPRSRPHGPNDRRSRNLG 120
 Qy 121 KVIDTLLTCGFADLGMGYIPLVGAPLGGAARALAHGVRLVEDGVNVATGNLPGCSFSIFLUA 180
 121 KVIDTLLTCGFADLGMGYIPLVGAPLGGAARALAHGVRLVEDGVNVATGNLPGCSFSIFLUA 180
 Db 181 LISCLTPASA 191
 Qy 181 LISCLTPASA 191
 Db 181 LISCLTPASA 191

RESULT 26

AAW12715
 ID AAW12715 standard; protein; 1006 AA.

XX AC AAW12715;
 XX DT 17-OCT-2003 (revised)
 DT 09-MAY-1997 (first entry)

DE HCV genome type 1a(H) amino acid residues 1-1006.
 XX HCV; influenza virus; vaccine; fusion protein; immunogen; core protein;
 KW NS1; plasmid PRIM14129.

XX OS Hepatitis C virus; genotype 1a(H).

XX FH Key
 FT Region
 FT 2. .166
 /label= "NS1
 /note= "amino acids 2-166 are utilised in a novel NS1-
 Core fusion protein"

XX PN W09701640-A2.
 XX PD 16-JAN-1997.
 PP 20-JUN-1996; 96WO-EP002764.
 XX PR 29-JUN-1995; 95GB-00013261.
 XX PA (SMIK) SSMITHKLINE BEECHAM BIOLOGICALS.

XX PI Cabezon Silva T, Momin PM, Garcon NMC;

XX DR WPI; 1997-100211/09.

XX PT New vaccines against hepatitis C virus - comprising Q521, 3D-MPL, an oil
 PT in water emulsion and a HCV core or envelope protein or immunogenic
 PT deriv.

XX PS Example 1; Page 15-16; 20pp; English.

XX CC Residues 1-1006 of the hepatitis C virus (HCV) genome type 1a(H) are
 CC shown in AAW12715. Residues 2-166 (from the core protein) have been

CC utilised in a novel fusion protein (see also AAW12714) with influenza
 CC virus NS1. The fusion protein can be formulated into novel vaccines
 CC against HCV. (Updated on 17-OCT-2003 to standardise OS field)

XX SQ Sequence 1006 AA;

Query Match 93.4%; Score 976; DB 2; Length 1006;
 Best Local Similarity 92.7%; Pred. No. 4 4.7e-83; 5; Mismatches 9; Indels 0; Gaps 0;
 Matches 177; Conservative

Qy 1 MSTLPKPKRKTKRNTNRPTDVKPFPGQQIVGGVYLPRRGPRGRGTRAKTSRSOPRG 60
 1 MSTLPKPKRKTKRNTNRPTDVKPFPGQQIVGGVYLPRRGPRGRGTRAKTSRSOPRG 60
 Db 61 RROPIPKARRQPOGRHWAQPGYPPWPLYGNBEGCGWAGWILSPRSRPHGPNDRRSRNLG 120
 Qy 61 RROPIPKARRQPOGRHWAQPGYPPWPLYGNBEGCGWAGWILSPRSRPHGPNDRRSRNLG 120
 61 RROPIPKARRQPOGRHWAQPGYPPWPLYGNBEGCGWAGWILSPRSRPHGPNDRRSRNLG 120
 Db 61 RROPIPKARRQPOGRHWAQPGYPPWPLYGNBEGCGWAGWILSPRSRPHGPNDRRSRNLG 120
 Qy 121 KVIDTLLTCGFADLGMGYIPLVGAPLGGAARALAHGVRLVEDGVNVATGNLPGCSFSIFLUA 180
 121 KVIDTLLTCGFADLGMGYIPLVGAPLGGAARALAHGVRLVEDGVNVATGNLPGCSFSIFLUA 180
 Db 181 LISCLTPASA 191
 Qy 181 LISCLTPASA 191
 Db 181 LISCLTPASA 191

RESULT 27

AAW12715
 ID AAR79221 standard; protein; 1648 AA.

XX AC AAR79221;
 XX DT 08-DEC-1995 (first entry)

XX DE PHCV176-encoded sequence.

XX KW PHCV176; amyloid precursor protein; APP; hepatitis C virus; HCV; E1; E2;
 KW fusion protein; HEK-293; vaccine; vector; prc/CMV; protein secretion;
 KW glycosylation.

XX OS Synthetic.

XX PN W09520664-A2.

XX PD 03-AUG-1995.

XX PT 27-JAN-1995; 95WO-US001087.

XX PR 28-JAN-1994; 94US-00186281.

XX PA (ABBO) ABBOTT LAB.

XX PI Watanaabe S, Yamaguchi J, Desai SM, Davare SG;
 XX DR WPI; 1995-275449/36.

XX PT New mammalian expression systems for HCV proteins - express fusion
 PT proteins comprising amyloid precursor protein and HCV E1 and/or E2
 PT protein.

XX PS Disclosure; Page 53-59; 89pp; English.

XX CC RNA from serum or plasma of a chimpanzee infected with HCV was converted
 CC to cDNA and PCR amplified using primers based on HCV sequences.⁷
 CC Adjacent HCV DNA fragments were generated which together encoded the HCV
 CC sequence given in AAR79232. Fragments from 2 clones, pICV141 (see AAR79222)
 CC and pICV150 (see AAR79223), were combined to generate PHCV176.
 CC This was used to construct APP-HCV E1+E2 fusion proteins that were
 CC glycosylated and secreted from HEK-293 transfecants

XX SQ Sequence 1648 AA;

The present sequence is Hepatitis C virus (HCV) H77C protein from chimpanzee #96A008 which lacks the hypervariable region one (HVR1) of HCV envelope 2 (E2). The HCV E2 protein lacking HVR1 (H77CHVR1) DNA is useful for producing infectious HCV and chimeric HCV viruses which are useful for identifying cell lines capable of supporting the replication of viruses. The infectious HCV and HVR1- chimeric HCV are used in the production of attenuated or inactivated vaccines which are useful for treating or preventing HCV in a mammal by immunisation. The host cells expressing the H77CHVR1 DNA is useful as an immunogen to stimulate a protective immune response to HCV. The immunogens are useful for producing protective antibodies to HCV. The antibodies produced are used in passive immunoprophylaxis for treatment of diseases caused by HCV in animals, especially humans. The H77CHVR1 DNA is also useful in gene therapy. Note: The present sequence is not found in the specification but is derived from Hepatitis C virus envelope protein 2 lacking hypervariable region 1 (HAE0042) referred as SEQ ID NO: 2 and shown in figure 1

Wed May 12 12:27:17 2004

us-09-084-691b-206.rag

Page 17

Search completed: May 12, 2004, 09:14:36
Job time : 60 secs

This Page Blank (uspto)

Run on: May 12, 2004, 09:16:06 ; Search time 41 Seconds
 (without alignments)
 1293.055 Million cell updates/sec

Copyright (c) 1993 - 2004 Compugen Ltd.

GenCore Version 5.1.6

OM protein - protein search, using sw model

Title: US-09-084-691B-206
 Sequence: 1 MSTLKPQPKRKRNTRRPT.....CSFSIPLLALISCLTPASA 191

Scoring table: BLOSUM62
 Gapop 10.0 , Gapext 0.5

Scored: 1140673 seqs, 277566755 residues

Total number of hits satisfying chosen parameters: 1140673

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
 Maximum Match 100%
 Listing first 90 summaries

Database : Published Applications AA.*

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep: *
 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep: *
 3: /cgn2_6/ptodata/1/pubpaa/US06_PUB.pep: *
 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep: *
 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep: *
 6: /cgn2_6/ptodata/1/pubpaa/PCTRS_PUBCOMB.pep: *
 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep: *
 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep: *
 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep: *
 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep: *
 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep: *
 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep: *
 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep: *
 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep: *
 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep: *
 16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep: *
 17: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep: *
 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match Length | DB ID | Description |
|------------|-------|--------------------|-------|--|
| 1 | 1045 | 100.0 | 319 | 12 US-10-651-165-228 Sequence 228, App Sequence 217, App Sequence 42, Appl |
| 2 | 980 | 93.8 | 12 | US-10-651-165-217 Sequence 229, App Sequence 181, App Sequence 23, Appl |
| 3 | 977 | 93.5 | 319 | 9 US-09-851-138-42 Sequence 2894, App Sequence 23, Appl |
| 4 | 977 | 93.5 | 319 | 12 US-10-651-165-229 Sequence 2894, App Sequence 23, Appl |
| 5 | 977 | 93.5 | 450 | 12 US-10-651-165-181 Sequence 2894, App Sequence 23, Appl |
| 6 | 977 | 93.5 | 2894 | 9 US-09-941-611-23 Sequence 2894, App Sequence 23, Appl |
| 7 | 977 | 93.5 | 2894 | 14 US-10-044-993-23 Sequence 2894, App Sequence 54, Appl |
| 8 | 976 | 93.4 | 2894 | 14 US-10-365-620-54 Sequence 2894, App Sequence 179, Appl |
| 9 | 976 | 93.4 | 450 | 12 US-10-651-165-179 Sequence 180, App Sequence 56, Appl |
| 10 | 976 | 93.4 | 450 | 12 US-10-651-165-180 Sequence 180, App Sequence 56, Appl |
| 11 | 976 | 93.4 | 473 | 15 US-10-345-620-56 Sequence 56, Appl |
| 12 | 976 | 93.4 | 3011 | 9 US-09-742-659-4 Sequence 4, Appl |
| 13 | 976 | 93.4 | 3011 | 9 US-09-952-572-9 Sequence 9, Appl |
| 14 | 976 | 93.4 | 3011 | 9 US-09-929-955-1 Sequence 1, Appl |
| 15 | 93.4 | 3011 | 9 | US-09-747-419-20 Sequence 20, Appl |

sequence 3, Appli
Sequence 14, Appli
Sequence 406, Appli
Sequence 1, Appli
Sequence 20, Appli
Sequence 3, Appli
Sequence 3, Appli
Sequence 2, Appli
Sequence 2, Appli
Sequence 2, Appli
Sequence 20, Appli
Sequence 20, Appli
Sequence 20, Appli
Sequence 190, Appli
Sequence 189, Appli
Sequence 1, Appli
Sequence 20, Appli
Sequence 219, Appli
Sequence 2, Appli
Sequence 3, Appli
Sequence 187, Appli
Sequence 193, Appli
Sequence 1, Appli
Sequence 48, Appli
Sequence 231, Appli
Sequence 7, Appli
Sequence 50, Appli
Sequence 50, Appli
Sequence 50, Appli
Sequence 50, Appli
Sequence 7, Appli
Sequence 11, Appli
Sequence 50, Appli
Sequence 50, Appli
Sequence 50, Appli
Sequence 53, Appli
Sequence 226, Appli
Sequence 50, Appli
Sequence 188, Appli
Sequence 55, Appli
Sequence 13, Appli
Sequence 50, Appli
Sequence 50, Appli
Sequence 194, Appli
Sequence 195, Appli
Sequence 218, Appli
Sequence 6, Appli
Sequence 52, Appli
Sequence 54, Appli
Sequence 3, Appli
Sequence 40, Appli
Sequence 191, Appli
Sequence 192, Appli
Sequence 9, Appli
Sequence 200, Appli
Sequence 13, Appli
Sequence 204, Appli
Sequence 144, Appli
Sequence 144, Appli
Sequence 144, Appli
Sequence 201, Appli
Sequence 2, Appli
Sequence 2, Appli
Sequence 8, Appli
Sequence 18, Appli
Sequence 206, Appli
Sequence 44, Appli
Sequence 230, Appli
Sequence 202, Appli
Sequence 58, Appli

89 915 87.6 459 15 US-10-365-620-60
90 895 85.6 319 12 US-10-651-165-211

Sequence 60, Appl
Sequence 211, App

ALIGNMENTS

RESULT 1

US-10-651-165-228

; Sequence 228, Application US/10651165

; Publication No. US20040047877A1

; GENERAL INFORMATION:

; APPLICANT: LEROUX-ROELS, Geert

; APPLICANT: DELEYS, Robert

; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPE OF HEPATITIS C

; FILE REFERENCE: 2551-94

; CURRENT APPLICATION NUMBER: US/10/651,165

; CURRENT FILING DATE: 2003-09-02

; PRIOR APPLICATION NUMBER: US/08/974,690C

; PRIOR FILING DATE: 1997-11-19

; PRIOR APPLICATION NUMBER: PCT/EP94/03555

; PRIOR FILING DATE: 1994-10-28

; PRIOR APPLICATION NUMBER: EP 93402718.6

; PRIOR FILING DATE: 1993-11-04

; NUMBER OF SEQ ID NOS: 286

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO: 228

; LENGTH: 319

; TYPE: PRT

; ORGANISM: hepatitis C virus

; US-10-651-165-228

Query Match Best Local Similarity 100.0%; Score 1045; DB 12; Length 319;

Matches 191; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 MSTLPKPKRKTKTENTRRPTDVKPGCGQTYGGVYLLPRRGPRLGVRATRKTSERSOPRG 60

Db 1 MSTNPKPKRKTKTENTRRPTDVKPGCGQTYGGVYLLPRRGPRLGVRATRKTSERSOPRG 60

Oy 61 RQQPIPKARQFGQRHAAQPGKPPWPLIGNEGGWAGMILSPRSRSPWGPNDPDRRSNLG 120

Db 61 RQQPIPKARQFGQRHAAQPGKPPWPLIGNEGGWAGMILSPRSRSPWGPNDPDRRSNLG 120

Oy 121 KVIDLTGCGPADLGMGYIPVVGAPGGVAALAHGVRAIEDGIVNATGNLPGCSFSIPLA 180

Db 121 KVIDLTGCGPADLGMGYIPVVGAPGGVAALAHGVRAIEDGIVNATGNLPGCSFSIPLA 180

Oy 181 LLSCLTTPASA 191

Db 181 LLSCLTTPASA 191

RESULT 3

US-09-851-138-42

; Sequence 42, Application US/09851138

; Publication No. US20020183508A1

; GENERAL INFORMATION:

; APPLICANT: MAERTENS, GHEERT

; STUYVER, LIEVEN

; TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGNOSTIC

; NUMBER OF SEQUENCES: 207

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: ARNOLD, WHITE & DURKEE

; STREET: P.O. BOX 4433

; CITY: HOUSTON

; STATE: TEXAS

; COUNTRY: USA

; ZIP: 77210-4433

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Microsoft Word 6.0 / ASCII text output

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/851,138

; FILING DATE: 09-May-2001

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/836,075

; FILING DATE: <Unknown>

; APPLICATION NUMBER: EP 94870166.9

; FILING DATE: 21 Oct 1994

; APPLICATION NUMBER: EP 95870076.7

; FILING DATE: 28 Jun 1995

; ATTORNEY/AGENT INFORMATION:

; NAME: KAMMERER, PATRICIA A

; REGISTRATION NUMBER: 29,775

; REFERENCE/DOCKET NUMBER: INNS:004

; INFORMATION FOR SEQ ID NO: 42:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 319 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

SEQUENCE DESCRIPTION: SEQ ID NO: 42:
US-09-851-138-42

Query Match 93.5%; Score 977; DB 9; Length 319;
Best Local Similarity 93.7%; Pred. No. 5e-79;
Matches 179; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

QY 1 MSTLPKPKRKTKRNTNRPPTDKPGGQIVCGVYLPRRGPRGLGVRAKRSERSOPRG 60
Db 1 MSTLPKPKRKTKRNTNRPMDVKPGGQIVCGVYLPRRGPRGLGVRAKRSERSOPRG 60

QY 61 RROPIPKARQPOQRHWAQPGYWPPLYGNEGCGWAGWILSPRSRPHGPNDRRSRNLG 120
Db 61 RROPIPKRQPTERSWGPQGPYWPPLYGNEGCGWAGWILSPRSRPNPNDPDRRSNLG 120

QY 121 KVIDTLTGCFADLIMGYIPFVVGAPLGVVAALAHGVRAEDGINYATGNLPGCSFSIHLA 180
Db 121 KVIDTLTGCFADLIMGYIPFVVGAPLGVVAALAHGVRAEDGINYATGNLPGCSFSIHLA 180

QY 181 LUSCLTPASA 191
Db 181 LUSCLTPASA 191

RESULT 4
US-10-165-165-229

; Sequence 229, Application US/10651165
; Publication No. US20040047877A1
; GENERAL INFORMATION:
; APPLICANT: DELEYS, Robert
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: MARTEENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPE OF HEPATITIS C
; FILE REFERENCE: 2551-94
; CURRENT APPLICATION NUMBER: US/10/651,165
; PRIORITY NUMBER: US/08/974,690C
; PRIORITY FILING DATE: 1997-11-19
; PRIORITY APPLICATION NUMBER: PCT/EP94/03555
; PRIORITY FILING DATE: 1994-10-28
; PRIORITY APPLICATION NUMBER: EP 93402718.6
; PRIORITY FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 181
; LENGTH: 450
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-10-651-165-181

Query Match 93.5%; Score 977; DB 12; Length 450;
Best Local Similarity 92.1%; Pred. No. 7.4e-79;
Matches 176; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

QY 1 MSTLPKPKRKTKRNTNRPPTDKPGGQIVCGVYLPRRGPRGLGVRAKRSERSOPRG 60
Db 1 MSTLPKPKRKTKRNTNRPMDVKPGGQIVCGVYLPRRGPRGLGVRAKRSERSOPRG 60

QY 61 RROPIPKARQPOQRHWAQPGYWPPLYGNEGCGWAGWILSPRSRPHGPNDRRSRNLG 120
Db 61 RROPIPKRQPTERSWGPQGPYWPPLYGNEGCGWAGWILSPRSRPNPNDPDRRSNLG 120

QY 121 KVIDTLTGCFADLIMGYIPFVVGAPLGVVAALAHGVRAEDGINYATGNLPGCSFSIHLA 180
Db 121 KVIDTLTGCFADLIMGYIPFVVGAPLGVVAALAHGVRAEDGINYATGNLPGCSFSIHLA 180

QY 181 LUSCLTPASA 191
Db 181 LUSCLTPASA 191

RESULT 5
US-10-651-165-181

; Sequence 181, Application US/10651165
; Publication No. US20040047877A1
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MARTEENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPE OF HEPATITIS C
; FILE REFERENCE: 2551-94
; CURRENT APPLICATION NUMBER: US/10/651,165
; PRIORITY APPLICATION NUMBER: US/08/974,690C
; PRIORITY FILING DATE: 1997-11-19
; PRIORITY APPLICATION NUMBER: PCT/EP94/03555
; PRIORITY FILING DATE: 1994-10-28
; PRIORITY APPLICATION NUMBER: EP 93402718.6
; PRIORITY FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 181
; LENGTH: 450
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-10-651-165-181

Query Match 93.5%; Score 977; DB 12; Length 450;

Best Local Similarity 92.1%; Pred. No. 7.4e-79;
Matches 176; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

QY 1 MSTLPKPKRKTKRNTNRPPTDKPGGQIVCGVYLPRRGPRGLGVRAKRSERSOPRG 60
Db 1 MSTLPKPKRKTKRNTNRPMDVKPGGQIVCGVYLPRRGPRGLGVRAKRSERSOPRG 60

QY 61 RROPIPKARQPOQRHWAQPGYWPPLYGNEGCGWAGWILSPRSRPHGPNDRRSRNLG 120
Db 61 RROPIPKRQPTERSWGPQGPYWPPLYGNEGCGWAGWILSPRSRPNPNDPDRRSNLG 120

QY 121 KVIDTLTGCFADLIMGYIPFVVGAPLGVVAALAHGVRAEDGINYATGNLPGCSFSIHLA 180
Db 121 KVIDTLTGCFADLIMGYIPFVVGAPLGVVAALAHGVRAEDGINYATGNLPGCSFSIHLA 180

QY 181 LUSCLTPASA 191
Db 181 LUSCLTPASA 191

RESULT 6
US-09-941-611-23

; Sequence 23, Application US/09941611
; Patent No. US20020106640A1
; GENERAL INFORMATION:

; APPLICANT: DELEYS, ROBERT J

; FOLLET, DIRK

; MARTENS, GEERT

; VAN HEVERSON, HUGO

; TITLE OF INVENTION: SYNTHETIC ANTIGENS FOR THE DETECTION OF

; NUMBER OF SEQUENCES: 23

; CORRESPONDENCE ADDRESS:

Query Match 93.5%; Score 977; DB 12; Length 319;
Best Local Similarity 93.7%; Pred. No. 5e-79;
Matches 179; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

QY 1 MSTLPKPKRKTKRNTNRPPTDKPGGQIVCGVYLPRRGPRGLGVRAKRSERSOPRG 60
Db 1 MSTLPKPKRKTKRNTNRPMDVKPGGQIVCGVYLPRRGPRGLGVRAKRSERSOPRG 60

QY 61 RROPIPKARQPOQRHWAQPGYWPPLYGNEGCGWAGWILSPRSRPHGPNDRRSRNLG 120
Db 61 RROPIPKRQPTERSWGPQGPYWPPLYGNEGCGWAGWILSPRSRPNPNDPDRRSNLG 120

Page 5

PRIOR FILING DATE: 2003-11-05
; PRIOR APPLICATION NUMBER: 60/3390,564
; PRIOR FILING DATE: 2003-06-20
; NUMBER OF SEQ ID NOS: 76
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO: 56
; LENGTH: 473
; TYPE: PRT
; ORGANISM: ORF of HCV core-TBD protein
; US-10-365-620-56

Query Match Best Local Similarity 93.4%; Score 976; DB 15; Length 473;
; Matches 177; Conservative 5; Mismatches 9; Indels 0; Gaps 0;
; SEQ ID NO: 1

QY 1 MSTLPKPKRKTKNTNRPTDVKPGGQIVGGVYLRRRGPKLGVRATKRSERSOPRG
Db 31 MSTNPKPORKTKNTNRPTDVKPGGQIVGGVYLRRRGPKLGVRATKRSERSOPRG

QY 61 RROPIPKARQPOGRHAWQPGVWPKPGQIVGGWAGWLSPRSRPHMCPNDPRRSNLG 60
Db 91 RROPIPKARRPEGTWQPGVWPKPGQIVGGWAGWLSPRSRPSWGPDTDRRSNLG 150

QY 121 KVIDLTGCFADLGYIPLVGAPLGGAARALAHGVRLVEDGVNVATGNLPGCSFSIFLLA 180
Db 151 KVIDLTGCFADLGYIPLVGAPLGGAARALAHGVRLVEDGVNVATGNLPGCSFSIFLLA 210

QY 181 LLSCLTPASA 191
Db 211 LLSCLTPASA 221

RESULT 12
US-09-742-659-4
; Sequence 4, Application US/09742659
; Patient No. US20010034019A1

GENERAL INFORMATION:
; APPLICANT: THONG, Zhi
; APPLICANT: Buckiewicz, Nancy J.
; APPLICANT: Zhong, Weidong
; APPLICANT: Ingravallo, Paul
; APPLICANT: Wright-Minogue, Jacqueline
; APPLICANT: Lau, Johnson Y.
; APPLICANT: Lemion, Stanley M.
; TITLE OF INVENTION: Chimeric HCV/GBV-B viruses
; FILE REFERENCE: ID0115
; CURRENT APPLICATION NUMBER: US/09/742,659
; PRIOR APPLICATION NUMBER: US/09/742,659
; PRIOR FILING DATE: 1999-12-22
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO: 4
; LENGTH: 3011
; TYPE: PRT
; ORGANISM: Hepatitis C virus
; US-09-742-659-4

Query Match Best Local Similarity 92.7%; Score 976; DB 9; Length 3011;
; Matches 177; Conservative 5; Mismatches 9; Indels 0; Gaps 0;
; SEQ ID NO: 1

RESULT 13
US-09-952-572-9
; Sequence 9, Application US/09952572
; Patient No. US2002011945A1

GENERAL INFORMATION:
; APPLICANT: HAWAII BIOTECHNOLOGY GROUP, Inc.
; APPLICANT: NAKANO, Eileen
; APPLICANT: CLEMENTS, David
; APPLICANT: HUMPHREYS, Tom
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITION OF HEPATITIS C
; FILE REFERENCE: HAWBIO110
; CURRENT APPLICATION NUMBER: US/09/952,572
; CURRENT FILING DATE: 2001-09-13
; PRIOR APPLICATION NUMBER: US 60/230,927
; PRIOR FILING DATE: 2000-09-13
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO: 9
; LENGTH: 3011
; TYPE: PRT
; ORGANISM: Hepatitis C Virus
; US-09-952-572-9

Query Match Best Local Similarity 92.7%; Score 976; DB 9; Length 3011;
; Matches 177; Conservative 5; Mismatches 9; Indels 0; Gaps 0;
; SEQ ID NO: 1

QY 1 MSTLPKPKRKTKNTNRPTDVKPGGQIVGGWAGWLSPRSRPHMCPNDPRRSNLG 60
Db 1 MSTNPKPORKTKNTNRPTDVKPGGQIVGGWAGWLSPRSRPSWGPDTDRRSNLG 60

QY 61 RROPIPKARQPOGRHAWQPGVWPKPGQIVGGWAGWLSPRSRPHMCPNDPRRSNLG 120
Db 61 RROPIPKARRPEGTWQPGVWPKPGQIVGGWAGWLSPRSRPSWGPDTDRRSNLG 120

QY 121 KVIDLTGCFADLGYIPLVGAPLGGAARALAHGVRLVEDGVNVATGNLPGCSFSIFLLA 180
Db 121 KVIDLTGCFADLGYIPLVGAPLGGAARALAHGVRLVEDGVNVATGNLPGCSFSIFLLA 180

QY 181 LLSCLTPASA 191
Db 181 LLSCLTPASA 191

RESULT 14
US-09-929-955-1
; Sequence 1, Application US/09929955
; Patient No. US2002016740A1

GENERAL INFORMATION:
; APPLICANT: Matti Sallberg
; APPLICANT: Catharina Holtgren
; TITLE OF INVENTION: VACCINES CONTAINING RIBAVIRIN AND
; TITLE OF INVENTION: METHODS OF USE THEREOF
; FILE REFERENCE: TR-PEP-23AU\$
; CURRENT APPLICATION NUMBER: US/09/929,955
; CURRENT FILING DATE: 2001-08-15
; PRIOR APPLICATION NUMBER: 09/705,547
; PRIOR FILING DATE: 2000-11-03
; PRIOR APPLICATION NUMBER: 60/229,175
; PRIOR FILING DATE: 2000-08-29
; PRIOR APPLICATION NUMBER: 60/225,767
; PRIOR FILING DATE: 2000-08-17
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: Fast-SEQ for Windows Version 4.0
; SEQ ID NO: 1
; LENGTH: 3011
; TYPE: PRT

Query Match Best Local Similarity 92.7%; Score 976; DB 9; Length 3011;
; Matches 177; Conservative 5; Mismatches 9; Indels 0; Gaps 0;
; SEQ ID NO: 1

QY 1 MSTLPKPKRKTKNTNRPTDVKPGGQIVGGWAGWLSPRSRPHMCPNDPRRSNLG 60
Db 1 MSTNPKPORKTKNTNRPTDVKPGGQIVGGWAGWLSPRSRPSWGPDTDRRSNLG 60

QY 61 RROPIPKARQPOGRHAWQPGVWPKPGQIVGGWAGWLSPRSRPHMCPNDPRRSNLG 120
Db 61 RROPIPKARRPEGTWQPGVWPKPGQIVGGWAGWLSPRSRPSWGPDTDRRSNLG 120

QY 121 KVIDLTGCFADLGYIPLVGAPLGGAARALAHGVRLVEDGVNVATGNLPGCSFSIFLLA 180
Db 121 KVIDLTGCFADLGYIPLVGAPLGGAARALAHGVRLVEDGVNVATGNLPGCSFSIFLLA 180

; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Hepatitis C virus sequence
; US-09-929-955-1

Query Match 93.4%; Score 976; DB 9; Length 3011;

Best Local Similarity 92.7%; Pred. No. 7.5e-78; Mismatches 5; Indels 0; Gaps 0;

Matches 177; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

Qy 1 MSTLPKPKRKTKNTNRPTDVKPGGGQIVGGVYLRRRGPRGTRAKTSERSOPRG 60

Db 1 MSTNPKPQPKTKNTNRPTDVKPGGGQIVGGVYLRRRGPRGTRAKTSERSOPRG 60

Qy 61 RROPIPKARPOQCRHWAOPGYWPPLYKNGCAGWMLSPRSRPHCPNDPDRRSRLG 120

Db 61 RROPIPKARRPEGRTRWQPGYWPPLYKNGCAGWMLSPRSRPSWGPDTDRRSRLG 120

Qy 121 KVIDLTLGCFADIMGYIPLVGAAPLGGAARALAHGVRLVEDGVNATGNLPGCSFSIILLA 180

Db 121 KVIDLTLGCFADIMGYIPLVGAAPLGGAARALAHGVRLVEDGVNATGNLPGCSFSIILLA 180

Qy 181 LLSCLTPASA 191

Db 181 LLSCLTPASA 191

RESULT 15

US-09-747-419-20

; Sequence 20, Application US/09747419

; Patent No. US20020155582A1

; GENERAL INFORMATION:

; APPLICANT: Lemon, Stanley

; TITLE OF INVENTION: REPLICATION COMPETENT HEPATITIS C VIRUS AND METHODS OF USE

; FILE REFERENCE: 265.007.0101

; CURRENT APPLICATION NUMBER: US/09/747.419

; CURRENT FILING DATE: 2000-12-23

; PRIOR APPLICATION NUMBER: US 60/171,909

; PRIOR FILING DATE: 1999-12-23

; NUMBER OF SEQ ID NOS: 34

; SOFTWARE: Patentin version 3.0

; SEQ ID NO: 20

; LENGTH: 3011

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Polyprotein

; -747-419-20

Query Match 93.4%; Score 976; DB 9; Length 3011;

Best Local Similarity 92.7%; Pred. No. 7.5e-78; Mismatches 5; Indels 0; Gaps 0;

Matches 177; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

Qy 1 MSTLPKPKRKTKNTNRPTDVKPGGGQIVGGVYLRRRGPRGTRAKTSERSOPRG 60

Db 1 MSTNPKPQPKTKNTNRPTDVKPGGGQIVGGVYLRRRGPRGTRAKTSERSOPRG 60

Qy 61 RROPIPKARPOQCRHWAOPGYWPPLYKNGCAGWMLSPRSRPHCPNDPDRRSRLG 120

Db 61 RROPIPKARRPEGRTRWQPGYWPPLYKNGCAGWMLSPRSRPSWGPDTDRRSRLG 120

Qy 121 KVIDLTLGCFADIMGYIPLVGAAPLGGAARALAHGVRLVEDGVNATGNLPGCSFSIILLA 180

Db 121 KVIDLTLGCFADIMGYIPLVGAAPLGGAARALAHGVRLVEDGVNATGNLPGCSFSIILLA 180

Qy 181 LLSCLTPASA 191

Db 181 LLSCLTPASA 191

RESULT 16

US-09-891-894-3

; Sequence 3, Application US/09891894

; Publication No. US20030013081A1

; GENERAL INFORMATION:

; APPLICANT: Olson, William

; APPLICANT: Maddon, Paul

; TITLE OF INVENTION: USES OF DC-SIGN AND DC-SIGNR FOR INHIBITING HEPATITIS C VIRUS INFI

; CURRENT APPLICATION NUMBER: US/09/891,894

; CURRENT FILING DATE: 2001-06-26

; NUMBER OF SEQ ID NOS: 3

; SOFTWARE: Patentin version 3.0

; SEQ ID NO: 3

; LENGTH: 3011

; TYPE: PRT

; ORGANISM: hepatitis c virus

; US-09-891-894-3

Query Match 93.4%; Score 976; DB 10; Length 3011;

Best Local Similarity 92.7%; Pred. No. 7.5e-78; Mismatches 5; Indels 0; Gaps 0;

Matches 177; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

Qy 1 MSTLPKPKRKTKNTNRPTDVKPGGGQIVGGVYLRRRGPRGTRAKTSERSOPRG 60

Db 1 MSTNPKPQPKTKNTNRPTDVKPGGGQIVGGVYLRRRGPRGTRAKTSERSOPRG 60

Qy 61 RROPIPKARPOQCRHWAOPGYWPPLYKNGCAGWMLSPRSRPHCPNDPDRRSRLG 120

Db 61 RROPIPKARRPEGRTRWQPGYWPPLYKNGCAGWMLSPRSRPSWGPDTDRRSRLG 120

Qy 121 KVIDLTLGCFADIMGYIPLVGAAPLGGAARALAHGVRLVEDGVNATGNLPGCSFSIILLA 180

Db 121 KVIDLTLGCFADIMGYIPLVGAAPLGGAARALAHGVRLVEDGVNATGNLPGCSFSIILLA 180

Qy 181 LLSCLTPASA 191

Db 181 LLSCLTPASA 191

RESULT 17

US-10-189-359-14

; Sequence 14, Application US/10189359

; Publication No. US20040039187A1

; GENERAL INFORMATION:

; APPLICANT: MARTIN, Annette

; APPLICANT: SANGER, DAVID V.

; TITLE OF INVENTION: Chimeric GB Virus B (GBV-B)

; FILE REFERENCE: USIG:2:8US

; CURRENT APPLICATION NUMBER: US/10/189,359

; CURRENT FILING DATE: 2002-07-03

; PRIOR APPLICATION NUMBER: 10/189,359

; PRIOR FILING DATE: 2002-07-03

; NUMBER OF SEQ ID NOS: 16

; SOFTWARE: Patentin Ver. 2.1

; SEQ ID NO: 14

; LENGTH: 3011

; TYPE: PRT

; ORGANISM: Hepatitis C virus

; US-10-189-359-14

Query Match 93.4%; Score 976; DB 12; Length 3011;

Best Local Similarity 92.7%; Pred. No. 7.5e-78; Mismatches 5; Indels 0; Gaps 0;

Matches 177; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

Qy 1 MSTLPKPKRKTKNTNRPTDVKPGGGQIVGGVYLRRRGPRGTRAKTSERSOPRG 60

Db 1 MSTNPKPQPKTKNTNRPTDVKPGGGQIVGGVYLRRRGPRGTRAKTSERSOPRG 60

Qy 61 RROPIPKARPOQCRHWAOPGYWPPLYKNGCAGWMLSPRSRPHCPNDPDRRSRLG 120

Db 61 RROPIPKARRPEGRTRWQPGYWPPLYKNGCAGWMLSPRSRPSWGPDTDRRSRLG 120

Qy 121 KVIDLTLGCFADIMGYIPLVGAAPLGGAARALAHGVRLVEDGVNATGNLPGCSFSIILLA 180

Db 121 KVIDLTLGCFADIMGYIPLVGAAPLGGAARALAHGVRLVEDGVNATGNLPGCSFSIILLA 180

Qy 181 LLSCLTPASA 191

Db 181 LLSCLTPASA 191

RESULT 18 ; FEATURE: OTHER INFORMATION: Hepatitis C virus sequence
; US-10-104-966-1
; Sequence 406, Application US/10296734
; Publication No. US2004054137A1
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; ATTORNEY: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; MEMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 406
; LENGTH: 3011
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC 1a consensus polyprotein
US-10-296-734-406

Query Match 93.4%; Score 976; DB 12; Length 3011;
Best Local Similarity 92.7%; Pred. No. 7.5e-78; Mismatches 9; Indels 0; Gaps 0;
Matches 177; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

Qy 1 MSTLPLPKRKTKRNTNRPTDVKPGGQIVGGVYLRRRGRLGVRAKTSERSOPRG 60
Db 1 MSTNPKPKRKTKRNTNRPTDVKPGGQIVGGVYLRRRGRLGVRAKTSERSOPRG 60

Qy 61 RROPIPKARQPOQRHWAQPGYWPMLYNGEGCGWAGWLSPRSRSPWGPNDPRRSNLG 120
Db 61 RROPIPKARRPEGRTRWQPGYWPMLYNGEGCWAGWLSPRSRSPWGPNDPRRSNLG 120

Qy 121 KVIDTLCGFADLMGYIPLVGAPLGGAARALAHGVRVLEDGVNYATGNLPGCSFS1FLLA 180
Db 121 KVIDTLCGFADLMGYIPLVGAPLGGAARALAHGVRVLEDGVNYATGNLPGCSFS1FLLA 180

Qy 181 LUSCLTVPASA 191
Db 181 LUSCLTVPASA 191

RESULT 20 ; FEATURE: OTHER INFORMATION: Hepatitis C virus sequence
; US-10-259-275-20
; Sequence 20, Application US/10259275
; Publication No. US20030125541A1
; GENERAL INFORMATION:
; APPLICANT: Lemon, Stanley M.
; ATTORNEY: Yi, MinKyung
; TITLE OF INVENTION: REPPLICATION COMPETENT HEPATITIS C VIRUS AND METHODS OF USE
; CURRENT APPLICATION NUMBER: US/10/259,275
; FILE REFERENCE: 205.0007.0120
; CURRENT FILING DATE: 2003-01-13
; PRIOR APPLICATION NUMBER: US 60/171,909
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: US 09/747,419
; PRIOR FILING DATE: 2000-12-23
; PRIOR APPLICATION NUMBER: US 60/325,236
; PRIOR FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: US 60/338,123
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 20
; LENGTH: 3011
; TYPE: PRT
; ORGANISM: artificial
; FEATURE:
; OTHER INFORMATION: Polyprotein
US-10-104-966-1

Query Match 93.4%; Score 976; DB 14; Length 3011;
Best Local Similarity 92.7%; Pred. No. 7.5e-78; Mismatches 9; Indels 0; Gaps 0;
Matches 177; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

Qy 1 MSTLPLPKRKTKRNTNRPTDVKPGGQIVGGVYLRRRGRLGVRAKTSERSOPRG 60
Db 1 MSTNPKPKRKTKRNTNRPTDVKPGGQIVGGVYLRRRGRLGVRAKTSERSOPRG 60

Qy 61 RROPIPKARQPOQRHWAQPGYWPMLYNGEGCGWAGWLSPRSRSPWGPNDPRRSNLG 120
Db 61 RROPIPKARRPEGRTRWQPGYWPMLYNGEGCWAGWLSPRSRSPWGPNDPRRSNLG 120

Qy 121 KVIDTLCGFADLMGYIPLVGAPLGGAARALAHGVRVLEDGVNYATGNLPGCSFS1FLLA 180
Db 121 KVIDTLCGFADLMGYIPLVGAPLGGAARALAHGVRVLEDGVNYATGNLPGCSFS1FLLA 180

Qy 181 LUSCLTVPASA 191
Db 181 LUSCLTVPASA 191

RESULT 21
US-10-184-150-3 Application US/10184150
; Sequence 3, Application US/10184150
; GENERAL INFORMATION:
; APPLICANT: Olson, William
; APPLICANT: Maddon, Paul
; TITLE OF INVENTION: USES OF DC-SIGN AND DC-SIGNR FOR INHIBITING HEPATITIS C VIRUS INF
; FILE REFERENCE: 2008/64396-A;JPW/MAF/DJK
; CURRENT APPLICATION NUMBER: US/10/184,150
; CURRENT FILING DATE: 2002-12-10
; PRIOR APPLICATION NUMBER: 09/891,894
; PRIOR FILING DATE: 2001-06-26
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO: 3
; LENGTH: 3011
; TYPE: PRT
; ORGANISM: Hepatitis C virus
; US-10-184-150-3

Query Match 93.4%; Score 976; DB 15; Length 3011;
Best Local Similarity 92.7%; Pred. No. 7; Sse-78;
Matches 177; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

Qy 1 MSTLPKPORKTKENTNRPPDKVKEPGQQIVGGYVLLPFRGPRGLGVRAKTSERSQRG 60
Db 61 RROPIPKRQPOCRHWAQPGYPWPLYAGEGCWAGWILSPRSRRPHMCNPDRRSRLG 120
Qy 121 KVDIPLCGFADLGMYIPVNGAPLGGAVALAHGVRAIEDGINVATGMLGCCFSIELA 180
Db 121 KVDIPLCGFADLGMYIPVNGAPLGGAVALAHGVRAIEDGINVATGMLGCCFSIELA 180
Qy 181 LUSCLTTPASA 191
Db 181 LUSCLTTPASA 191

RESULT 23
US-09-238-076-2 Application US/09238076
; Sequence 2, Application US/09238076
; Patient No. US200201025401
; GENERAL INFORMATION:
; APPLICANT: RICE, CHARLES et al.
; TITLE OF INVENTION: FUNCTIONAL DNA CLONE FOR HEPATITIS C
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HOWELL & HAERKAMP, L.C.
; STREET: 7733 FORSYTH BLVD., SUITE 1400
; CITY: ST. LOUIS
; STATE: MO
; COUNTRY: USA
; ZIP: 63105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/238,076
; FILING DATE:
; CLASSIFICATION:
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 09/034,756
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: HOLLAND, DONALD R.
; REGISTRATION NUMBER: 35,197
; REFERENCE DOCKET NUMBER: 6039-4831
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 314-727-5188
; TELEFAX: 314-727-6092
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3012 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; FRAGMENT TYPE: N-terminal
; US-09-238-076-2

Query Match 93.4%; Score 976; DB 9; Length 3012;
Best Local Similarity 92.7%; Pred. No. 7; Sse-78;
Matches 177; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

Qy 1 MSTLPKPORKTKENTNRPPDKVKEPGQQIVGGYVLLPFRGPRGLGVRAKTSERSQRG 60
Db 61 RROPIPKRQPOCRHWAQPGYPWPLYAGEGCWAGWILSPRSRRPHMCNPDRRSRLG 120
Qy 61 RROPIPKRQPOCRHWAQPGYPWPLYAGEGCWAGWILSPRSRRPHMCNPDRRSRLG 120
Db 61 RROPIPKRQPOCRHWAQPGYPWPLYAGEGCWAGWILSPRSRRPHMCNPDRRSRLG 120

RESULT 24
 US-09-995-937-2
 Publication No. US20030028010A1
 Sequence 2, Application US/09995937
 ; Sequence 2, Application US/09995937
 GENERAL INFORMATION:
 APPLICANT: RICE, CHARLES et al.
 TITLE OF INVENTION: FUNCTIONAL DNA CLONE FOR HEPATITIS C VIRUS (HCV) AND USES THEREOF
 NUMBER OF SEQUENCES: 21
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: HOWELL & HAERKAMP, L.C.
 STREET: 773 FORSYTH BLVD., SUITE 1400
 CITY: ST. LOUIS
 STATE: MO
 COUNTRY: USA
 ZIP: 63105
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/995, 937
 FILING DATE: 28-NO-2001
 CLASSIFICATION: <Unknown>
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US/09/034, 756
 FILING DATE: 04-MAY-1998
 ATTORNEY/AGENT INFORMATION:
 NAME: HOLLAND, DONALD R.
 REGISTRATION NUMBER: 35,197
 REFERENCE/DOCKET NUMBER: 6029-4831
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 314-727-5188
 TELEFAX: 314-727-6092
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 3012 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 MOLECULE TYPE: protein
 HYPOTHETICAL: NO
 FRAGMENT TYPE: N-terminal
 SEQUENCE DESCRIPTION: SEQ ID NO: 2:
 US-09-995-937-2
 Query Match 93.4%; Score 976; DB 10; Length 3012;
 Best Local Similarity 92.7%; Pred. No. 7.5e-78;
 Matches 177; Conservative 5; Mismatches 9; Indels 0; Gaps 0;
 Qy 1 MSTLPKPORKTKRNTNRRTDKPKEPGGQIVGGVYLPRRGPRGLGRVTRKTSRSOPRG 60
 Db 1 MSTNPKPORKTKRNTNRRTDKPKEPGGQIVGGVYLPRRGPRGLGRVTRKTSRSOPRG 60
 Qy 61 RROPIKARQOGRHWAQPGYWPWLYGNNGGWAGWLSPRSRSPHRGPNDRRSRNLG 120
 Db 61 RROPIKARQOGRHWAQPGYWPWLYGNNGGWAGWLSPRSRSPHRGPNDRRSRNLG 120
 Qy 121 KVIDLTGCFADLGMYIPVVGAPLGVAALAHGVRAIEDGIVATGNLPGCSFSIFILA 180
 Db 121 KVIDLTGCFADLGMYIPVVGAPLGVAALAHGVRAIEDGIVATGNLPGCSFSIFILA 180
 Qy 181 LLSCLTVPASA 191
 Db 181 LLSCLTVPASA 191
 181 LLSCLTVPASA 191
 RESULT 25
 US-09-917-563-2
 Publication No. US20030073080A1
 Sequence 2, Application US/09917563
 GENERAL INFORMATION:
 APPLICANT: RICE, CHARLES et al.
 TITLE OF INVENTION: FUNCTIONAL DNA CLONE FOR HEPATITIS C VIRUS (HCV) AND USES THEREOF
 NUMBER OF SEQUENCES: 21
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: HOWELL & HAERKAMP, L.C.
 STREET: 773 FORSYTH BLVD., SUITE 1400
 CITY: ST. LOUIS
 STATE: MO
 COUNTRY: USA
 ZIP: 63105
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/238, 076
 FILING DATE: 26-JAN-1999
 ATTORNEY/AGENT INFORMATION:
 NAME: HOLLAND, DONALD R.
 REGISTRATION NUMBER: 35,197
 REFERENCE/DOCKET NUMBER: 6029-4831
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 314-727-5188
 TELEFAX: 314-727-6092
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 3012 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 MOLECULE TYPE: protein
 HYPOTHETICAL: NO
 FRAGMENT TYPE: N-terminal
 SEQUENCE DESCRIPTION: SEQ ID NO: 2:
 US-09-917-563-2
 Query Match 93.4%; Score 976; DB 10; Length 3012;
 Best Local Similarity 92.7%; Pred. No. 7.5e-78;
 Matches 177; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

US-09-238-076-20
; Sequence 20, Application US/09238076
; Patient No. US2002010254041
; GENERAL INFORMATION:
; APPLICANT: RICE, CHARLES et al.
; TITLE OF INVENTION: FUNCTIONAL DNA CLONE FOR HEPATITIS C
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HOWELL & HAERKAMP, L.C.
; STREET: 7733 FORSYTH BLVD., SUITE 1400
; CITY: ST. LOUIS
; STATE: MO
; COUNTRY: USA
; ZIP: 63105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/238,076
; FILING DATE:
; CLASSIFICATION:
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 09/034,756
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: HOLLAND, DONALD R.
; REGISTRATION NUMBER: 35,197
; REFERENCE/DOCKET NUMBER: 6029-4831
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 314-727-5188
; TELEFAX: 314-727-6092
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3011 amino acids
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOLETICAL: NO
; FRAGMENT TYPE: N-terminal
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOLETICAL: NO
; FRAGMENT TYPE: N-terminal
; SEQUENCE DESCRIPTION: SEQ ID NO: 20:
; US-09-995-937-20
; Query Match 93.0%; Score 972; DB 9; Length 3011;
; Best Local Similarity 92.1%; Pred. No. 1.7e-77;
; Matches 176; Conservative 6; Mismatches 9; Indels 0; Gaps 0;
; QY |||||||C|H|S|L|P|K|R|Q|R|K|T|K|N|R|P|T|D|V|K|P|G|G|O|I|V|G|Y|V|L|P|R|G|P|R|L|G|V|R|A|T|K|S|E|R|S|Q|R|G 60
; QY |||||||C|H|S|L|P|K|R|Q|R|K|T|K|N|R|P|T|D|V|K|P|G|G|O|I|V|G|Y|V|L|P|R|G|P|R|L|G|V|R|A|T|K|S|E|R|S|Q|R|G 60
; Db 1 MSTNPKPORKTKRNTNRPRQDVEFPGGQIVGGVYLIPRRGPRLGVRAKTSERSQRG 60
; Qy 61 RROPIPKARQPRQRHWAGPGYPPFLYQEGCGWAGWILSPRSRPHCPNDRRSNLG 120
; Qy |||||||C|H|S|L|P|K|R|Q|R|K|T|K|N|R|P|T|D|V|K|P|G|G|O|I|V|G|Y|V|L|P|R|G|P|R|L|G|V|R|A|T|K|S|E|R|S|Q|R|G 120
; Db 61 RROPIPKARQPRQRHWAGPGYPPFLYQEGCGWAGWILSPRSRPHCPNDRRSNLG 120
; Qy 121 KVIDLTGCFADLMGYIPVVGAPLGGAVALAHGVAILEDGINYATGNGPGCSFSIFLLA 180
; Qy |||||||C|H|S|L|P|K|R|Q|R|K|T|K|N|R|P|T|D|V|K|P|G|G|O|I|V|G|Y|V|L|P|R|G|P|R|L|G|V|R|A|T|K|S|E|R|S|Q|R|G 180
; Db 121 KVIDLTGCFADLMGYIPVVGAPLGGAVALAHGVAILEDGINYATGNGPGCSFSIFLLA 180
; Qy 181 LLSCLTPASA 191
; Db 181 LLSCLTPASA 191
; RESULT 27
; US-09-995-937-20
; Sequence 20, Application US/09995937
; Publication No. US20030073080A1
; GENERAL INFORMATION:
; APPLICANT: RICE, CHARLES et al.
; TITLE OF INVENTION: FUNCTIONAL DNA CLONE FOR HEPATITIS C
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HOWELL & HAERKAMP, L.C.
; STREET: 7733 FORSYTH BLVD., SUITE 1400
; CITY: ST. LOUIS
; STATE: MO
; COUNTRY: USA
; ZIP: 63105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/995,937
; FILING DATE: 28-No. US20030028010A1-2001
; CLASSIFICATION: <Unknown>
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US/09/034,756
; FILING DATE: 04-May-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: HOLLAND, DONALD R.
; REGISTRATION NUMBER: 35,197
; REFERENCE/DOCKET NUMBER: 6029-4831
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 314-727-5188
; TELEFAX: 314-727-6092
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3011 amino acids
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOLETICAL: NO
; FRAGMENT TYPE: N-terminal
; SEQUENCE DESCRIPTION: SEQ ID NO: 20:
; US-09-995-937-20
; Query Match 93.0%; Score 972; DB 10; Length 3011;
; Best Local Similarity 92.1%; Pred. No. 1.7e-77;
; Matches 176; Conservative 6; Mismatches 9; Indels 0; Gaps 0;
; QY |||||||C|H|S|L|P|K|R|Q|R|K|T|K|N|R|P|T|D|V|K|P|G|G|O|I|V|G|Y|V|L|P|R|G|P|R|L|G|V|R|A|T|K|S|E|R|S|Q|R|G 60
; QY |||||||C|H|S|L|P|K|R|Q|R|K|T|K|N|R|P|T|D|V|K|P|G|G|O|I|V|G|Y|V|L|P|R|G|P|R|L|G|V|R|A|T|K|S|E|R|S|Q|R|G 60
; Do 1 MSTNPKPORKTKRNTNRPRQDVEFPGGQIVGGVYLIPRRGPRLGVRAKTSERSQRG 60
; Qy 61 RROPIPKARQPRQRHWAGPGYPPFLYQEGCGWAGWILSPRSRPHCPNDRRSNLG 120
; Qy |||||||C|H|S|L|P|K|R|Q|R|K|T|K|N|R|P|T|D|V|K|P|G|G|O|I|V|G|Y|V|L|P|R|G|P|R|L|G|V|R|A|T|K|S|E|R|S|Q|R|G 120
; Do 61 RROPIPKARQPRQRHWAGPGYPPFLYQEGCGWAGWILSPRSRPHCPNDRRSNLG 120
; Qy 121 KVIDLTGCFADLMGYIPVVGAPLGGAVALAHGVAILEDGINYATGNGPGCSFSIFLLA 180
; Qy |||||||C|H|S|L|P|K|R|Q|R|K|T|K|N|R|P|T|D|V|K|P|G|G|O|I|V|G|Y|V|L|P|R|G|P|R|L|G|V|R|A|T|K|S|E|R|S|Q|R|G 180
; Do 121 KVIDLTGCFADLMGYIPVVGAPLGGAVALAHGVAILEDGINYATGNGPGCSFSIFLLA 180
; Qy 181 LLSCLTPASA 191
; Do 181 LLSCLTPASA 191

CITY: ST. LOUIS
 STATE: MO
 COUNTRY: USA
 ZIP: 63105

COMPUTER READABLE FORM:
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, version #1.30

CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/917,563
 FILING DATE: 27-Jul-2001
 CLASSIFICATION: <Unknown>
 PRIORITY APPLICATION DATA:
 APPLICATION NUMBER: 09/238,076
 FILING DATE: 26-JAN-1999
 ATTORNEY/AGENT INFORMATION:
 NAME: HOLLAND, DONALD R.
 REGISTRATION NUMBER: 35,197
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 314-727-5188
 TELEFAX: 314-727-6092

INFORMATION FOR SEQ ID NO: 20:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 3011 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 HYPOTHETICAL: NO
 FRAGMENT TYPE: N-terminal
 FRAGMENT DESCRIPTION: SEQ ID NO: 20:

US-09-917-563-20

Query Match
 Best Local Similarity 93.0%; Score 972; DB 10; Length 3011;
 Matches 176; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

Oy 1 MSTLPKQPKTKNTNRPTDKFPGGQIVGVVYLPRRGRLGTRAKTTSERSOPRG 60
 Db 1 MSTNPKPQPKTKNTNRPTDKFPGGQIVGVVYLPRRGRLGTRAKTTSERSOPRG 60

Oy 61 RRQPIPKARQPOGRHQAQPGYPPWLYNGNECGWAGWLSPRSRSPWPGPDRRRSLNG 120
 Db 61 RRQPIPKARRGRTWAQPGYPPWLYNGEFGWAGWLSPRSRSPWPGPDRRRSLNG 120

Oy 121 KVIDLTGCFADLMGYIIPVVGAPLGGAVALAHGVRAEDGINAYATGNLPGCSFSIPLA 180
 Db 121 KVIDLTGCFADLMGYIIPVVGAPLGGAVALAHGVRAEDGINAYATGNLPGCSFSIPLA 180

Oy 181 LLSCLTTPASA 191
 Db 181 LLSCLTTPASA 191

RESULT 30
 US-10-651-165-189
 Sequence 189, Application US/10651165
 Publication No. US20040047877A1

GENERAL INFORMATION:
 APPLICANT: LEROUX-ROELS, Geert
 APPLICANT: DELES, Robert
 APPLICANT: MAERTENS, Geert
 TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPE OF HEPATITIS C
 FILE REFERENCE: 2551-94

CURRENT APPLICATION NUMBER: US/10/651,165
 CURRENT FILING DATE: 2003-09-02
 PRIORITY APPLICATION NUMBER: US/08/974,690C
 PRIORITY FILING DATE: 1997-11-19
 PRIORITY APPLICATION NUMBER: PCT/EP94/03555
 PRIORITY FILING DATE: 1994-10-28

PRIORITY APPLICATION NUMBER: EP 93402718.6
 PRIORITY FILING DATE: 1993-11-04
 NUMBER OF SEQ ID NOS: 286
 SOFTWARE: PatentIn version 3.1
 SEQ ID NO 189
 LENGTH: 450
 TYPE: PRT
 ORGANISM: hepatitis C virus

US-10-651-165-189

RESULT 29
 US-10-651-165-190
 Sequence 190, Application US/10651165
 Publication No. US20040047877A1

GENERAL INFORMATION:
 APPLICANT: LEROUX-ROELS, Geert
 APPLICANT: DELES, Robert
 APPLICANT: MAERTENS, Geert
 TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPE OF HEPATITIS C
 TITLE OF INVENTION: VIRUS
 FILE REFERENCE: 2551-94
 CURRENT APPLICATION NUMBER: US/10/651,165
 CURRENT FILING DATE: 2003-09-02
 PRIORITY APPLICATION NUMBER: US/08/974,690C
 PRIORITY FILING DATE: 1997-11-19
 PRIORITY APPLICATION NUMBER: PCT/EP94/03555
 PRIORITY FILING DATE: 1994-10-28
 PRIORITY APPLICATION NUMBER: EP 93402718.6
 PRIORITY FILING DATE: 1993-11-04
 NUMBER OF SEQ ID NOS: 286
 SOFTWARE: PatentIn version 3.1
 SEQ ID NO 189
 LENGTH: 450
 TYPE: PRT
 ORGANISM: hepatitis C virus

US-10-651-165-189

Query Match
 Best Local Similarity 92.7%; Score 969; DB 12; Length 450;
 Matches 177; Conservative 4; Mismatches 10; Indels 0; Gaps 0;

Oy 1 MSTLPKQPKTKNTNRPTDKFPGGQIVGVVYLPRRGRLGTRAKTTSERSOPRG 60
 Db 1 MSTNPKPQPKTKNTNRPTDKFPGGQIVGVVYLPRRGRLGTRAKTTSERSOPRG 60

Oy 61 RRQPIPKARQPOGRHQAQPGYPPWLYNGNECGWAGWLSPRSRSPWPGPDRRRSLNG 120
 Db 61 RRQPIPKARRGRTWAQPGYPPWLYNGEFGWAGWLSPRSRSPWPGPDRRRSLNG 120

Oy 121 KVIDLTGCFADLMGYIIPVVGAPLGGAVALAHGVRAEDGINAYATGNLPGCSFSIPLA 180
 Db 121 KVIDLTGCFADLMGYIIPVVGAPLGGAVALAHGVRAEDGINAYATGNLPGCSFSIPLA 180

/ Wed May 12 12:27:17 2004

us-09-084-691b-206.rapb

Page 13

Oy 181 LUSCLTTPASA 191
| | | | |
Db 181 LUSCLTIPASA 191

Search completed: May 12, 2004, 09:21:46
Job time : 43 secs

This Page Blank (uspto)

APPLICANT: BURKH, J., MILLER, R. H. AND
 TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
 TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
 TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
 TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
 NUMBER OF SEQUNCE: 263
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: MORGAN & FINNEGAN
 STREET: 345 PARK AVENUE
 CITY: NEW YORK
 STATE: NEW YORK
 COUNTRY: USA
 ZIP: 10154

COMPUTER READABLE FORM:
 COMPUTER: IBM PC COMPATIBLE
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: WORDPERFECT 5.1
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/290,665A
 FILING DATE: 15-AUG-1994
 CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
 NAME: RICHARD W. BORK
 REGISTRATION NUMBER: 36,459
 REFERENCE/DOCKET NUMBER: 2026-4116
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 758-4800
 TELEFAX: (212) 751-6849
 TELEX: 421792

INFORMATION FOR SEQ ID NO: 206:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 191 amino acids
 STRANDEDNESS: unknown
 TOPOLOGY: unknown
 ORIGINAL SOURCE:
 ORGANISM: homosapiens
 INDIVIDUAL ISOLATE: HK2
 US-08-290-665A-06

Query Match 100.0%; Score 1045; DB 2; Length 191;
 Best Local Similarity 100.0%; Pred. No. 7.4e-97; Mismatches 0; Indels 0; Gaps 0;
 Matches 191; Conservative 0; MisMatches 0; Indels 0; Gaps 0;

```

Q 1 MSTLIPKPKRTKNTNRPTDVKPGG3QIVG3VYLLPRRGRLGVTRTKRSERSOPRG 60
Db 1 MSTLIPKPKRTKNTNRPTDVKPGG3QIVG3VYLLPRRGRLGVTRTKRSERSOPRG 60
Qy 61 RROPIPKARQPOGRHWAQPGYPPRGLGEGCWAGWILLSPRSRPHMCPNDRRSNLG 120
Db 61 RROPIPKARQPOGRHWAQPGYPPRGLGEGCWAGWILLSPRSRPHMCPNDRRSNLG 120
Qy 61 RROPIPKARQPOGRHWAQPGYPPRGLGEGCWAGWILLSPRSRPHMCPNDRRSNLG 120
Db 61 RROPIPKARQPOGRHWAQPGYPPRGLGEGCWAGWILLSPRSRPHMCPNDRRSNLG 120
Qy 121 KVIDLTLTCGFADIMGYIPVVGAPLGVAALAHGVRAIEDGINYATGNLPGCSFSIFLA 180
Db 121 KVIDLTLTCGFADIMGYIPVVGAPLGVAALAHGVRAIEDGINYATGNLPGCSFSIFLA 180
Qy 181 LISCLTPASA 191
Db 181 LISCLTPASA 191

```

RESULT 2

PCT-US95-10398-206

Sequence 206, Application PC/TUS9510398
 GENERAL INFORMATION:
 APPLICANT: BURKH, J., MILLER, R. H. AND
 APPLICANT: PURCELL, R. H.
 TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
 TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
 TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
 TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
 NUMBER OF SEQUNCE: 263
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: MORGAN & FINNEGAN
 STREET: 345 PARK AVENUE
 CITY: NEW YORK
 STATE: NEW YORK
 COUNTRY: USA
 ZIP: 10154

TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
 TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
 NUMBER OF SEQUNCE: 263
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: MORGAN & FINNEGAN
 STREET: 345 PARK AVENUE
 CITY: NEW YORK
 STATE: NEW YORK
 COUNTRY: USA
 ZIP: 10154

COMPUTER READABLE FORM:
 COMPUTER: IBM PC COMPATIBLE
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: WORDPERFECT 5.1
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US95/10398
 FILING DATE: 15-AUG-1995
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/086,428
 FILING DATE: 29-JUNE-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/290/665
 FILING DATE: 15-AUGUST-1994
 ATTORNEY/AGENT INFORMATION:
 NAME: RICHARD W. BORK
 REGISTRATION NUMBER: 36,459
 REFERENCE/DOCKET NUMBER: 2026-4116
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 758-4800
 TELEFAX: (212) 751-6849
 TELEX: 421792

INFORMATION FOR SEQ ID NO: 206:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 191 amino acids
 STRANDEDNESS: unknown
 TOPOLOGY: unknown
 ORIGINAL SOURCE:
 ORGANISM: homosapiens
 INDIVIDUAL ISOLATE: HK2
 PCT-US95-10398-206

Query Match 100.0%; Score 1045; DB 5; Length 191;
 Best Local Similarity 100.0%; Pred. No. 7.4e-97; Mismatches 0; Indels 0; Gaps 0;
 Matches 191; Conservative 0; MisMatches 0; Indels 0; Gaps 0;

```

Q 1 MSTLIPKPKRTKNTNRPTDVKPGG3QIVG3VYLLPRRGRLGVTRTKRSERSOPRG 60
Db 1 MSTLIPKPKRTKNTNRPTDVKPGG3QIVG3VYLLPRRGRLGVTRTKRSERSOPRG 60
Qy 61 RROPIPKARQPOGRHWAQPGYPPRGLGEGCWAGWILLSPRSRPHMCPNDRRSNLG 120
Db 61 RROPIPKARQPOGRHWAQPGYPPRGLGEGCWAGWILLSPRSRPHMCPNDRRSNLG 120
Qy 61 RROPIPKARQPOGRHWAQPGYPPRGLGEGCWAGWILLSPRSRPHMCPNDRRSNLG 120
Db 61 RROPIPKARQPOGRHWAQPGYPPRGLGEGCWAGWILLSPRSRPHMCPNDRRSNLG 120
Qy 121 KVIDLTLTCGFADIMGYIPVVGAPLGVAALAHGVRAIEDGINYATGNLPGCSFSIFLA 180
Db 121 KVIDLTLTCGFADIMGYIPVVGAPLGVAALAHGVRAIEDGINYATGNLPGCSFSIFLA 180
Qy 181 LISCLTPASA 191
Db 181 LISCLTPASA 191

```

RESULT 3

US-08-635-866C-228

Sequence 228, Application US/08635886C
 GENERAL INFORMATION:
 APPLICANT: LEROUW ROELS, Geert
 APPLICANT: DELEYNS, Robert
 APPLICANT: MAERTENS, Geert

RESULT 6
PCT-US95-10398-191
; Sequence 191, Application PC/TUS9510398
; GENERAL INFORMATION:
; APPLICANT: PURCELL, R.H.
; COMPUTER: BURKH, J., MILLER, R.H. AND
; TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
; TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
; TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
; TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
; TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
; NUMBER OF SEQUENCES: 263
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/10398
; FILING DATE: 15-AUG-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/086,428
; FILING DATE: 29 JUNE 1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/290/665
; FILING DATE: 15 AUGUST 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: RICHARD W. BORK
; REGISTRATION NUMBER: 36,459
; REFERENCE/DOCKET NUMBER: 2026-4116
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEX: (212) 751-6849
; INFORMATION FOR SEQ ID NO: 191:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 191 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; ORIGINAL SOURCE:
; ORGANISM: homosapiens
; INDIVIDUAL ISOLATE: Z4
; PCT-US95-10398-191

Query Match 95.3%; Score 996; DB 5; Length 191;
Best Local Similarity 94.8%; Pred. No. 5.9e-92; 6; Indels 0; Gaps 0;
Matches 181; Conservative 4; Mismatches 6;

Qy 1 MSTNLPKPKRTKENTNRPTDVKPGGQIVGVVYLRRGPPLGVRAKRSERSQPG 60
Db 1 MSTNLPKPKRTKENTNRPTDVKPGGQIVGVVYLRRGPPLGVRAKRSERSQPG 60
Qy 61 RQPIPKARQPRQRHWAQPGYPWPLYGNBGGWAGWLSPRSRPHQCPNDRRSRNIG 120
Db 61 RQPIPKARQPRQRHWAQPGYPWPLYGNBGGWAGWLSPRSRPHQCPNDRRSRNIG 120
Qy 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAALAHGVRAEDGINYATGNLPGCSFSIPLA 180
Db 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAALAHGVRAEDGINYATGNLPGCSFSIPLA 180
Qy 181 LLSCLTTPASA 191
; Sequence 191, Application PC/TUS9510398
; GENERAL INFORMATION:
; APPLICANT: PURCELL, R.H.
; COMPUTER: BURKH, J., MILLER, R.H. AND
; TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
; TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
; TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
; TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
; TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
; NUMBER OF SEQUENCES: 263
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/290,665A
; FILING DATE: 15-AUG-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: RICHARD W. BORK
; REGISTRATION NUMBER: 36,459
; REFERENCE/DOCKET NUMBER: 2026-4116
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEX: (212) 751-6849
; INFORMATION FOR SEQ ID NO: 192:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 191 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; ORIGINAL SOURCE:
; ORGANISM: homosapiens
; INDIVIDUAL ISOLATE: Z8
; US-08-290-665A-192
; Query Match 94.0%; Score 982; DB 2; length 191;
; Best Local Similarity 93.7%; Pred. No. 1.5e-90; 5; Mismatches 7; Indels 0; Gaps 0;
; Matches 179; Conservative 5;

Qy 1 MSTNLPKPKRTKENTNRPTDVKPGGQIVGVVYLRRGPPLGVRAKRSERSQPG 60
Db 1 MSTNLPKPKRTKENTNRPTDVKPGGQIVGVVYLRRGPPLGVRAKRSERSQPG 60
Qy 61 RQPIPKARQPRQRHWAQPGYPWPLYGNBGGWAGWLSPRSRPHQCPNDRRSRNIG 120
Db 61 RQPIPKARQPRQRHWAQPGYPWPLYGNBGGWAGWLSPRSRPHQCPNDRRSRNIG 120
Qy 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAALAHGVRAEDGINYATGNLPGCSFSIPLA 180
Db 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAALAHGVRAEDGINYATGNLPGCSFSIPLA 180
Qy 181 LLSCLTTPASA 191

Db 181 LSLCLTVBASA 191

RESULT 8

US-08-290-665A-193

Sequence 193, Application US/08290665A

Patent No. 582852

GENERAL INFORMATION:

APPLICANT: BURKH, J., MILLER, R.H. AND

APPLICANT: PURCELL, R.H.

TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED

TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND

TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS

TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE

TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE

TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES

NUMBER OF SEQUENCES: 263

CORRESPONDENCE ADDRESS:

ADDRESSEE: MORGAN & FINNEGAN

STREET: 345 PARK AVENUE

STATE: NEW YORK

COUNTRY: USA

ZIP: 10154

COMPUTER READABLE FORM:

MEDIUM TYPE: FLOPPY DISK

COMPUTER: IBM PC COMPATIBLE

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: WORDPERFECT 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/290,665A

FILING DATE: 15-AUG-1994

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: RICHARD W. BORK

REGISTRATION NUMBER: 36,459

REFERENCE/DOCKET NUMBER: 2026-4116

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 758-4800

TELEFAX: (212) 751-6849

TELEX: 421792

INFORMATION FOR SEQ ID NO: 193:

SEQUENCE CHARACTERISTICS:

LENGTH: 191 amino acids

TYPE: amino acid

STRANDEDNESS: unknown

TOPOLOGY: unknown

ORIGINAL SOURCE:

ORGANISM: homosapiens

INDIVIDUAL ISOLATE: Z1

US-08-290-665A-193

RESULT 9

PCT-US95-10398-192

Sequence 192, Application PC/TUS9510398

GENERAL INFORMATION:

APPLICANT: BURKH, J., MILLER, R.H. AND

APPLICANT: PURCELL, R.H.

TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED

TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND

TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS

TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE

TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES

NUMBER OF SEQUENCES: 263

CORRESPONDENCE ADDRESS:

ADDRESSEE: MORGAN & FINNEGAN

STREET: 345 PARK AVENUE

CITY: NEW YORK

STATE: NEW YORK

COUNTRY: USA

ZIP: 10154

COMPUTER READABLE FORM:

MEDIUM TYPE: FLOPPY DISK

COMPUTER: IBM PC COMPATIBLE

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: WORDPERFECT 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US95/10398

FILING DATE: 15-AUG-1995

CLASSIFICATION:

PRIORITY APPLICATION DATA:

APPLICATION NUMBER: 08/086,428

FILING DATE: 29 JUNE 1993

PRIORITY APPLICATION DATA:

APPLICATION NUMBER: 08/290/665

FILING DATE: 15 AUGUST 1994

ATTORNEY/AGENT INFORMATION:

NAME: RICHARD W. BORK

REGISTRATION NUMBER: 36,459

REFERENCE/DOCKET NUMBER: 2026-4116

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 758-4800

TELEFAX: (212) 751-6849

TELEX: 421792

INFORMATION FOR SEQ ID NO: 192:

SEQUENCE CHARACTERISTICS:

LENGTH: 191 amino acids

TYPE: amino acid

STRANDEDNESS: unknown

TOPOLOGY: unknown

ORIGINAL SOURCE:

ORGANISM: homosapiens

INDIVIDUAL ISOLATE: Z8

US-08-290-665A-192

RESULT 8

US-08-290-665A-193

Sequence 193, Application US/08290665A

Patent No. 582852

GENERAL INFORMATION:

APPLICANT: BURKH, J., MILLER, R.H. AND

APPLICANT: PURCELL, R.H.

TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED

TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND

TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS

TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE

TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES

NUMBER OF SEQUENCES: 263

CORRESPONDENCE ADDRESS:

ADDRESSEE: MORGAN & FINNEGAN

STREET: 345 PARK AVENUE

CITY: NEW YORK

STATE: NEW YORK

COUNTRY: USA

ZIP: 10154

COMPUTER READABLE FORM:

MEDIUM TYPE: FLOPPY DISK

COMPUTER: IBM PC COMPATIBLE

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: WORDPERFECT 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US95/10398

FILING DATE: 15-AUG-1995

CLASSIFICATION:

PRIORITY APPLICATION DATA:

APPLICATION NUMBER: 08/086,428

FILING DATE: 29 JUNE 1993

PRIORITY APPLICATION DATA:

APPLICATION NUMBER: 08/290/665

FILING DATE: 15 AUGUST 1994

ATTORNEY/AGENT INFORMATION:

NAME: RICHARD W. BORK

REGISTRATION NUMBER: 36,459

REFERENCE/DOCKET NUMBER: 2026-4116

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 758-4800

TELEFAX: (212) 751-6849

TELEX: 421792

INFORMATION FOR SEQ ID NO: 193:

SEQUENCE CHARACTERISTICS:

LENGTH: 191 amino acids

TYPE: amino acid

STRANDEDNESS: unknown

TOPOLOGY: unknown

ORIGINAL SOURCE:

ORGANISM: homosapiens

INDIVIDUAL ISOLATE: Z1

US-08-290-665A-193

RESULT 9

US-09-084-691b-206.rai

Sequence 192, Application PC/TUS9510398

General Information:

Applicant: BURKH, J., MILLER, R.H. AND

Applicant: PURCELL, R.H.

Title of Invention: Nucleotide and deduced

Title of Invention: amino acid sequences of the envelope 1 and

Title of Invention: core genes of isolates of hepatitis c virus

Title of Invention: and the use of reagents derived from these

Title of Invention: sequences in diagnostic methods and vaccines

Number of Sequences: 263

Correspondence Address:

Addressee: MORGAN & FINNEGAN

Street: 345 PARK AVENUE

City: NEW YORK

State: NEW YORK

Country: USA

Zip: 10154

Computer Readable Form:

Medium Type: FLOPPY DISK

Computer: IBM PC COMPATIBLE

Operating System: PC-DOS/MS-DOS

Software: WORDPERFECT 5.1

Current Application Data:

Application Number: PCT/US95/10398

Filing Date: 15-AUG-1995

Classification:

Priority Application Data:

Application Number: 08/086,428

Filing Date: 29 JUNE 1993

Prior Application Data:

Application Number: 08/290/665

Filing Date: 15 AUGUST 1994

Attorney/Agent Information:

Name: RICHARD W. BORK

Registration Number: 36,459

Reference/Docket Number: 2026-4116

Telecommunication Information:

Telephone: (212) 758-4800

Telefax: (212) 751-6849

Telex: 421792

Information for Seq ID No: 192:

Sequence Characteristics:

Length: 191 amino acids

Type: amino acid

Strandedness: unknown

Topology: unknown

Original Source:

Organism: homosapiens

Individual Isolate: Z8

US-08-290-665A-192

Query Match

Best Local Similarity 94.0%; Score 982; DB 2; Length 191;

Matches 179; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Matches 179; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Matches 179; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Matches 179; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Matches 179; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Matches 179; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Matches 179; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Matches 179; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Matches 179; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Matches 179; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Matches 179; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Matches 179; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Matches 179; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Matches 179; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Matches 179; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Matches 179; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Matches 179; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Matches 179; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Matches 179; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Matches 179; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Matches 179; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Matches 179; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Matches 179; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Matches 179; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Matches 179; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Matches 179; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Matches 179; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Matches 179; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Matches 179; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Matches 179; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Matches 179; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Matches 179; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Matches 179; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Matches 179; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Matches 179; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Matches 179; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Matches 179; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Matches 179; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Matches 179; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Matches 179; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Matches 179; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Matches 179; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Matches 179; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Matches 179; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Matches 179; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Matches 179; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Matches 179; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Matches 179; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Matches 179; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Matches 179; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Matches 179; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Matches 179; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Matches 179; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Matches 179; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Matches 179; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Matches 179; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Matches 179; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Matches 179; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Matches 179; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Matches 179; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Matches 179; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Matches 179; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Query Match

RESULT 10

PCT-US95-10398-193

Sequence 193, Application PC/TUS9510398

GENERAL INFORMATION:

APPLICANT: PURCELL, R.H.

TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED

TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND

TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS

TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE

NUMBER OF SEQUENCES: 263

CORRESPONDENCE ADDRESS:

ADDRESSEE: MORGAN & FINNEGAN

STREET: 345 PARK AVENUE

CITY: NEW YORK

STATE: NEW YORK

COUNTRY: USA

ZIP: 10154

COMPUTER READABLE FORM:

MEDIUM TYPE: FLOPPY DISK

COMPUTER: IBM PC COMPATIBLE

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: WORDPERFECT 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US95/10398

FILING DATE: 15 AUG-1995

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/086,428

FILING DATE: 29 JUNE 1993

PRIORITY APPLICATION DATA:

APPLICATION NUMBER: 08/290/665

FILING DATE: 15 AUGUST 1994

ATTORNEY/AGENT INFORMATION:

NAME: RICHARD W. BORK

REGISTRATION NUMBER: 36,459

REFERENCE/DOCKET NUMBER: 2026-4116

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 758-4800

TELEFAX: (212) 75-6849

TELEPAX: 421792

INFORMATION FOR SEQ ID NO: 193:

SEQUENCE CHARACTERISTICS:

LENGTH: 191 amino acids

TYPE: amino acid

STRANDEDNESS: unknown

TOPOLOGY: unknown

ORIGINAL SOURCE:

ORGANISM: homosapiens

INDIVIDUAL ISOLATE: ZL

PCT-US95-10398-193

Query Match 94.0%; Score 982; DB 5; Length 191;
Best Local Similarity 93.7%; Pred. No. 1.5e-90; Mismatches 7; Indels 0; Gaps 0;
Matches 179; ConservativeQy 1 MSTLPKPQRTKTENTNRPTDKFPGGGQIVGGVYLRLRGPGLGVRAKTRKTSERSQPRG
Db 1 MSTNPKPQRTKTENTNRPTDKFPGGGQIVGGVYLRLRGPGLGVRAKTRKTSERSQPRG 60Qy 61 RROPIPKARQPGHWAQPGYWPFLYQWPGNCGWAGWLSPRSRSPWNGPDRRRNLG 120
Db 61 RROPIPKARQPGHWAQPGYWPFLYQWPGNCGWAGWLSPRSRSPWNGPDRRRNLG 120Qy 121 KVIDTLLQGFAIDLGYIPVGAPLGGLGVAAALANGVRAEDGINYATGMLPGSFSFILLA 180
Db 121 KVIDTLLQGFAIDLGYIPVGAPLGGLGVAAALANGVRAEDGINYATGMLPGSFSFILLA 180Qy 181 LLSCLTPASA 191
Db 181 LLSCLTPASA 191

181 LLSCLTPASA 191

RESULT 11

US-08-290-665A-172

Sequence 172, Application US/08290665A

GENERAL INFORMATION:

APPLICANT: BURKH, J., MILLER, R.H. AND

TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED

TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND

TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS

TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE

NUMBER OF SEQUENCES: 263

CORRESPONDENCE ADDRESS:

ADDRESSEE: MORGAN & FINNEGAN

STREET: 345 PARK AVENUE

CITY: NEW YORK

STATE: NEW YORK

COUNTRY: USA

ZIP: 10154

COMPUTER READABLE FORM:

MEDIUM TYPE: FLOPPY DISK

COMPUTER: IBM PC COMPATIBLE

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: WORDPERFECT 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/290,665A

FILING DATE: 15-AUG-1994

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: RICHARD W. BORK

REGISTRATION NUMBER: 36,459

REFERENCE/DOCKET NUMBER: 2026-4116

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 758-4800

TELEFAX: (212) 75-6849

TELEPAX: 421792

INFORMATION FOR SEQ ID NO: 172:

SEQUENCE CHARACTERISTICS:

LENGTH: 191 amino acids

TYPE: amino acid

STRANDEDNESS: unknown

TOPOLOGY: unknown

ORIGINAL SOURCE:

ORGANISM: homosapiens

INDIVIDUAL ISOLATE: HK3

US-08-290-665A-172

Query Match 93.9%; Score 981; DB 2; Length 191;

Best Local Similarity 93.7%; Pred. No. 1.9e-90; Mismatches 7; Indels 0; Gaps 0;

Matches 179; Conservative

Qy 1 MSTLPKPQRTKTENTNRPTDKFPGGGQIVGGVYLRLRGPGLGVRAKTRKTSERSQPRG 60
Db 1 MSTNPKPQRTKTENTNRPTDKFPGGGQIVGGVYLRLRGPGLGVRAKTRKTSERSQPRG 60Qy 61 RROPIPKARQPGHWAQPGYWPFLYQWPGNCGWAGWLSPRSRSPWNGPDRRRNLG 120
Db 61 RROPIPKARQPGHWAQPGYWPFLYQWPGNCGWAGWLSPRSRSPWNGPDRRRNLG 120Qy 121 KVIDTLLQGFAIDLGYIPVGAPLGGLGVAAALANGVRAEDGINYATGMLPGSFSFILLA 180
Db 121 KVIDTLLQGFAIDLGYIPVGAPLGGLGVAAALANGVRAEDGINYATGMLPGSFSFILLA 180Qy 181 LLSCLTPASA 191
Db 181 LLSCLTPASA 191

181 LLSCLTPASA 191

Sequence 172, Application PC/TUS9510398
 GENERAL INFORMATION:
 APPLICANT: BUKH, J., MILLER, R.H. AND

APPLICANT: PURCELL, R.H.

TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED

TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS

TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE

NUMBER OF SEQUENCES: 263

CORRESPONDENCE ADDRESS:

ADDRESSEE: MORGAN & FINNEGAN

STREET: 345 PARK AVENUE

CITY: NEW YORK

STATE: NEW YORK

COUNTRY: USA

ZIP: 10154

COMPUTER READABLE FORM:

MEDIUM TYPE: FLOPPY DISK

COMPUTER: IBM PC COMPATIBLE

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: WORDPERFECT 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US95/10398

FILING DATE: 15-AUG-1995

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/086,428

FILING DATE: 29 JUNE 1993

PRIOR ATTORNEY/AGENT INFORMATION:

NAME: RICHARD W. BORK

REGISTRATION NUMBER: 36,459

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/290/665

FILING DATE: 15-AUG-1994

PRIOR ATTORNEY/AGENT INFORMATION:

NAME: RICHARD W. BORK

REGISTRATION NUMBER: 2026-4116

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 758-4800

TELEFAX: (212) 751-6849

TELEX: 421792

INFORMATION FOR SEQ ID NO: 172:

SEQUENCE CHARACTERISTICS:

LENGTH: 191 amino acids

TYPE: amino acid

STRANDEDNESS: unknown

TOPOLOGY: unknown

ORIGINAL SOURCE:

ORGANISM: homosapiens

INDIVIDUAL ISOLATE: HK3

ORIGINAL SOURCE:

INDIVIDUAL ISOLATE: HK3

us-08-290-665A-197
 Sequence 197, Application US/08290665A
 ; Patent No. 5842852

GENERAL INFORMATION:
 APPLICANT: BUKH, J., MILLER, R.H. AND

APPLICANT: PURCELL, R.H.

TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED

TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND

TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS

TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE

TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES

NUMBER OF SEQUENCES: 263

CORRESPONDENCE ADDRESS:

ADDRESSEE: MORGAN & FINNEGAN

STREET: 345 PARK AVENUE

CITY: NEW YORK

STATE: NEW YORK

COUNTRY: USA

ZIP: 10154

COMPUTER READABLE FORM:

MEDIUM TYPE: FLOPPY DISK

COMPUTER: IBM PC COMPATIBLE

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: WORDPERFECT 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/290,665A

FILING DATE: 15-AUG-1994

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: RICHARD W. BORK

REGISTRATION NUMBER: 36,459

PRIOR APPLICATION NUMBER: 2026-4116

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 758-4800

TELEFAX: (212) 751-6849

TELEX: 421792

INFORMATION FOR SEQ ID NO: 197:

SEQUENCE CHARACTERISTICS:

LENGTH: 191 amino acids

TYPE: amino acid

STRANDEDNESS: unknown

TOPOLOGY: unknown

ORIGINAL SOURCE:

ORGANISM: homosapiens

INDIVIDUAL ISOLATE: DK13

Us-08-290-665A-197
 Query Match 93.8%; Score 980; DB 2; Length 191;
 Best Local Similarity 93.7%; Pred. No. 2.3e-90; Indels 0; Gaps 0;
 Matches 179; Conservative 4; Mismatches 8;

Query 1 MSTLPKPKRKTKRNTNRPTDKPFGGGQIVGGVYLRRRGRGRGTRAKTKRSRPG

DB 1 MSTNPKPKRKTKRNTNRPTDKPFGGGQIVGGVYLRRRGRGRGTRAKTKRSRPG

Query 61 RQDPIPKARQPCRHRWAGPGYPMPLYNGEGCAGWGLLSPRSRPHCPNDRRRSRLG

DB 61 RQDPIPKARQPCRHRWAGPGYPMPLYNGEGCAGWGLLSPRSRPHCPNDRRRSRLG

Query 121 KVIDLTQGFAIDLGYIIVVVGAPLGGVAAALAHGVRAIEDGINYATGMLPGCSFSFLA

DB 121 KVIDLTQGFAIDLGYIIVVVGAPLGGVAAALAHGVRAIEDGINYATGMLPGCSFSFLA

Query 181 LSLCLTPASA 191

DB 181 LSLCLTPASA 191

APPLICANT: PURCELL, R.H.
 TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
 TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
 TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
 TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
 TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
 NUMBER OF SEQUENCES: 263
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: MORGAN & FINNEGAN
 STREET: 345 PARK AVENUE
 CITY: NEW YORK
 STATE: NEW YORK
 COUNTRY: USA
 ZIP: 10154
 COMPUTER READABLE FORM:
 MEDIUM TYPE: FLOPPY DISK
 COMPUTER: IBM PC COMPATIBLE
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: WORDPERFECT 5.1
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US95/10398
 FILING DATE: 15-AUG-1995
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/086,428
 FILING DATE: 29 JUNE 1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/290/665
 FILING DATE: 15 AUGUST 1994
 ATTORNEY/AGENT INFORMATION:
 NAME: RICHARD W. BORK
 REGISTRATION NUMBER: 36,459
 REFERENCE/DOCKET NUMBER: 2026-4116
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 755-4800
 TELEX: 421792
 INFORMATION FOR SEQ ID NO: 197:
 SEQENCE CHARACTERISTICS:
 LENGTH: 191 amino acids
 TYPE: amino acid
 STRANDEDNESS: unknown
 TOPOLOGY: unknown
 ORIGINAL SOURCE:
 ORGANISM: homosapiens
 INDIVIDUAL ISOLATE: DK13
 PCT-US95-10398-197

Query Match 93.8%; Score 980; DB 5; Length 191;
 Best Local Similarity 93.7%; Pred. No. 4.4e-90; Gaps 0;
 Matches 179; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY |||||P|K|R|Q|R|K|T|G|T|N|R|R|P|D|V|K|F|P|G|G|Q|I|V|G|Y|L|L|R|R|P|R|G|R|I|G|R|A|T|R|K|S|E|R|P|R|G
 1 MSTLPKQRKTGENTNRPTDVKFPGGQIVGGYLRLRRPGLGRATRKTSERSOPRG 60
 DB |||||P|K|R|Q|R|K|T|G|T|N|R|R|P|D|V|K|F|P|G|G|Q|I|V|G|Y|L|L|R|R|P|R|G|R|I|G|R|A|T|R|K|S|E|R|P|R|G
 1 MSTNPKQRKTENRPTDVKFPGGQIVGGYLRLRRPGLGRATRKTSERSOPRG 60

QY |||||R|Q|R|P|K|R|Q|R|K|T|G|T|N|R|R|P|D|V|K|F|P|G|G|Q|I|V|G|Y|L|L|R|R|P|R|G|R|I|G|R|A|T|R|K|S|E|R|P|R|G
 61 RRQTPKQRQPOCRHMAQPGYWPMLPVNGECWAGWNLLSPRSRSPRHPGPNDPRRSRNLG 120
 DB |||||R|Q|R|P|K|R|Q|R|K|T|G|T|N|R|R|P|D|V|K|F|P|G|G|Q|I|V|G|Y|L|L|R|R|P|R|G|R|I|G|R|A|T|R|K|S|E|R|P|R|G
 61 RRQTPKQRQPOCRHMAQPGYWPMLPVNGECWAGWNLLSPRSRSPRHPGPNDPRRSRNLG 120

QY |||||K|V|D|T|L|C|G|F|A|D|I|G|Y|I|P|V|V|G|A|P|V|G|G|A|P|A|L|A|H|G|V|A|L|E|D|G|Y|A|T|G|N|L|P|G|C|F|S|I|F|L|A
 121 KVIDLTGCFADIGYIPVVGAAPVGGAparalAHGValeDGyATGNLPGCSFSIFLIA 180
 DB |||||K|V|D|T|L|C|G|F|A|D|I|G|Y|I|P|V|V|G|A|P|V|G|G|A|P|A|L|A|H|G|V|A|L|E|D|G|Y|A|T|G|N|L|P|G|C|F|S|I|F|L|A
 121 KVIDLTGCFADIGYIPVVGAAPVGGAparalAHGValeDGyATGNLPGCSFSIFLIA 180

QY |||||L|L|S|C|L|T|P|A|S|A| 191
 181 LLSCLTTPASA 191
 DB |||||L|L|S|C|L|T|V|P|A|S|A| 191
 181 LLSCLTVPASA 191

RESULT 16
 US-08-974-690C-217
 Sequence 217, Application US/08974690C
 ; Patent No. 6613333
 ; GENERAL INFORMATION:
 ; APPLICANT: LEROUX-ROELS, Geert
 ; APPLICANT: DELEY, Robert
 ; APPLICANT: MARTENS, Geert
 ; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPE OF HEPATITIS C
 ; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPE OF HEPATITIS C
 ; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPE OF HEPATITIS C
 ; CURRENT APPLICATION NUMBER: US/08/974,690C
 ; CURRENT FILING DATE: 1997-11-19
 ; PRIOR APPLICATION NUMBER: PCT/EP94/03555
 ; PRIOR FILING DATE: 1994-10-28
 ; PRIOR APPLICATION NUMBER: EP 93402718.6
 ; PRIOR FILING DATE: 1993-11-04
 ; NUMBER OF SEQ ID NOS: 286
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO: 217
 ; LENGTH: 319
 ; TYPE: PRT
 ; ORGANISM: hepatitis C virus
 US-08-974-690C-217

Query Match 93.8%; Score 980; DB 4; Length 319;
 Best Local Similarity 93.7%; Pred. No. 4.4e-90; Gaps 0;
 Matches 179; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY |||||M|S|T|L|P|K|R|Q|R|K|T|G|T|N|R|R|P|D|V|K|F|P|G|G|Q|I|V|G|Y|L|L|R|R|P|R|G|R|I|G|R|A|T|R|K|S|E|R|P|R|G
 1 MSTLPKQRKTENRPTDVKFPGGQIVGGYLRLRRPGLGRATRKTSERSOPRG 60

QY |||||R|O|P|I|P|K|R|Q|R|K|T|G|T|N|R|R|P|D|V|K|F|P|G|G|Q|I|V|G|Y|L|L|R|R|P|R|G|R|I|G|R|A|T|R|K|S|E|R|P|R|G
 61 RROPIPKARQPOCRHMAQPGYWPMLPVNGECWAGWNLLSPRSRSPRHPGPNDPRRSRNLG 120
 DB |||||R|O|P|I|P|K|R|Q|R|K|T|G|T|N|R|R|P|D|V|K|F|P|G|G|Q|I|V|G|Y|L|L|R|R|P|R|G|R|I|G|R|A|T|R|K|S|E|R|P|R|G
 61 RROPIPKARQPOCRHMAQPGYWPMLPVNGECWAGWNLLSPRSRSPRHPGPNDPRRSRNLG 120

QY |||||K|V|D|T|L|C|G|F|A|D|I|G|Y|I|P|V|V|G|A|P|V|G|G|A|P|A|L|A|H|G|V|A|L|E|D|G|Y|A|T|G|N|L|P|G|C|F|S|I|F|L|A
 121 KVIDLTGCFADIGYIPVVGAAPVGGAparalAHGValeDGyATGNLPGCSFSIFLIA 180
 DB |||||K|V|D|T|L|C|G|F|A|D|I|G|Y|I|P|V|V|G|A|P|V|G|G|A|P|A|L|A|H|G|V|A|L|E|D|G|Y|A|T|G|N|L|P|G|C|F|S|I|F|L|A
 121 KVIDLTGCFADIGYIPVVGAAPVGGAparalAHGValeDGyATGNLPGCSFSIFLIA 180

QY |||||L|L|S|C|L|T|P|A|S|A| 191
 181 LLSCLTTPASA 191
 DB |||||L|L|S|C|L|T|V|P|A|S|A| 191

RESULT 15
 US-08-635-886C-217
 Sequence 217, Application US/08635886C
 ; Patent No. 6555114

Db 1 MSTNPKPQKRTKNTNRPMDVKPGGGQIVGCVYUPLRRGRGLGVATRKTSERSOPRG 60
 Qy 61 RQQPPIKARQPGHAWQPGYVPMPLYNEGCCWAGWLLSPRSRSPNGPNDRRSRNLG 120
 Db 61 RQQPPIKARQPGHAWQPGYVPMPLYNEGCCWAGWLLSPRSRSPNGPNDRRSRNLG 120
 Db 121 KVIDTLCGFAIDLMLGYIPVVGAPLGGAALAHGVRAIEDGINYATGNLPGCSFSIPLLA 180
 Qy 121 KVIDTLCGFAIDLMLGYIPVVGAPVGVARALAHGVRAVEDGINYATGNLPGCSFSIPLLA 180
 Qy 181 LLSCLTPASA 191
 Db 181 LLSCLTPASA 191
 ;
RESULT 17
 US-08-290-665A-195
 Sequence 195, Application US/08290665A
 Patent No. 582852
 GENERAL INFORMATION:
 APPLICANT: PURCELL, R.H.
 TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
 TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
 TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
 TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
 TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
 NUMBER OF SEQUENCES: 263
 CORRESPONDENCE ADDRESS:
 ADDRESSEES: MORGAN & FINNEGAN
 STREET: 345 PARK AVENUE
 CITY: NEW YORK
 STATE: NY
 COUNTRY: USA
 ZIP: 10154
 COMPUTER READABLE FORM:
 MEDIUM TYPE: FLOPPY DISK
 COMPUTER: IBM PC COMPATIBLE
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: WORDPERFECT 5.1
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/290,665A
 FILING DATE: 15-AUG-1994
 ATTORNEY/AGENT INFORMATION:
 NAME: RICHARD W. BORK
 REGISTRATION NUMBER: 36,459
 REFERENCE/DOCKET NUMBER: 2026-4116
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 758-4800
 TELEX: 421792
 INFORMATION FOR SEQ ID NO: 195:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 191 amino acids
 TYPE: amino acid
 STRANDEDNESS: unknown
 TOPOLOGY: unknown
 ORIGINAL SOURCE:
 ORGANISM: homosapiens
 INDIVIDUAL ISOLATE: Z6
 ;
 US-08-290-665A-195
 Query Match 93.6%; Score 978; DB 2; Length 191;
 Best Local Similarity 93.2%; Pred. No. 3; 7e-90;
 Matches 178; Conservative 5; Mismatches 8; Indels 0; Gaps 0;
 ;
 Qy 1 MSTLPKPQKRTKNTNRPMDVKPGGGQIVGCVYUPLRRGRGLGVATRKTSERSOPRG 60
 Db 1 MSTLPKPQKRTKNTNRPMDVKPGGGQIVGCVYUPLRRGRGLGVATRKTSERSOPRG 60
 Qy 61 RQQPPIKARQPGHAWQPGYVPMPLYNEGCCWAGWLLSPRSRSPNGPNDRRSRNLG 120
 ;
 PCT-US95-10398-195
 Query Match 93.6%; Score 978; DB 5; Length 191;
 Best Local Similarity 93.2%; Pred. No. 3; 7e-90;
 Matches 178; Conservative 5; Mismatches 8; Indels 0; Gaps 0;
 ;
 Qy 1 MSTLPKPQKRTKNTNRPMDVKPGGGQIVGCVYUPLRRGRGLGVATRKTSERSOPRG 60
 Db 1 MSTLPKPQKRTKNTNRPMDVKPGGGQIVGCVYUPLRRGRGLGVATRKTSERSOPRG 60
 Qy 61 RQQPPIKARQPGHAWQPGYVPMPLYNEGCCWAGWLLSPRSRSPNGPNDRRSRNLG 120

RESULT 19
US-08-290-665A-196
Sequence 196 Application US/08290665A.
; GENERAL INFORMATION:
; APPLICANT: BURKH, J., MILLER, R.H. AND
; APPLICANT: PURCELL, R.H.
TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
NUMBER OF SEQUENCES: 263
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/290,665A
FILING DATE: 15-AUG-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: RICHARD W. BORK
REGISTRATION NUMBER: 36,459
REFERENCE DOCKET NUMBER: 2026-4116
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 196:
SEQUENCE CHARACTERISTICS:
LENGTH: 191 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
ORIGINAL SOURCE:
ORGANISM: homosapiens
INDIVIDUAL ISOLATE: 27
US-08-290665A-196

Query Match 93.5%; Score 977; DB 2; Length 191;
Best Local Similarity 93.2%; Pred. No. 4.7e-90; 8; Indels 0; Gaps 0;
Matches 178; Conservative 5; Mismatches 8;

QY 1 MSTLPKPKRKTKRTRNRPDKFPGGQVGGVILPRRGPRGRATRKTSERSOPRG 60
QY 1 MSTLPKPKRKTKRTRNRPDKFPGGQVGGVILPRRGPRGRATRKTSERSOPRG 60
QY 61 RROPIPKARQPOGRHWAQPGYPWPLYGNEGCGWAGWLSPRSRSPHRWGPNDPRRSRNLG 120
QY 61 RRQPIPKARRSEGRGWAQCPYWPMLYGNEGCGWAGWLSPRSRSPHRWGPNDPRRSRNLG 120
QY 121 KVDTLTCGFDLGYIPVGAPLGCGVAALAHGTRATEIGINTYATGNLGCFSIFLLA 180

RESULT 20
PCT-US95-10398-196
Sequence 196 Application PCT/TU9510398
; GENERAL INFORMATION:
; APPLICANT: BURKH, J., MILLER, R.H. AND
; APPLICANT: PURCELL, R.H.
TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
NUMBER OF SEQUENCES: 263
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/10398
FILING DATE: 15-AUG-1995
CLASSIFICATION:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 08/086,428
FILING DATE: 29 JUNE 1993
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 08/290,665
FILING DATE: 15 AUGUST 1994
ATTORNEY/AGENT INFORMATION:
NAME: RICHARD W. BORK
REGISTRATION NUMBER: 36,459
REFERENCE DOCKET NUMBER: 2026-4116
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 196:
SEQUENCE CHARACTERISTICS:
LENGTH: 191 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
ORIGINAL SOURCE:
ORGANISM: homosapiens
INDIVIDUAL ISOLATE: 27
PCT-US95-10398-196

Query Match 93.5%; Score 977; DB 5; Length 191;
Best Local Similarity 93.2%; Pred. No. 4.7e-90; 8; Indels 0; Gaps 0;
Matches 178; Conservative 5; Mismatches 8;

QY 1 MSTLPKPKRKTKRTRNRPDKFPGGQVGGVILPRRGPRGRATRKTSERSOPRG 60
QY 1 MSTLPKPKRKTKRTRNRPDKFPGGQVGGVILPRRGPRGRATRKTSERSOPRG 60
QY 61 RROPIPKARQPOGRHWAQPGYPWPLYGNEGCGWAGWLSPRSRSPHRWGPNDPRRSRNLG 120
QY 61 RRQPIPKARRSEGRGWAQCPYWPMLYGNEGCGWAGWLSPRSRSPHRWGPNDPRRSRNLG 120
QY 61 RROPIPKARRSEGRSWAQPGYPWMLYGNEGCGWAGWLSPRSRSPHRWGPNDPRRSRNLG 120

QY 121 KVIDLTTCGFAIDLGYIPVVGAPLGGAALAHGVRAIEDGINYATGNLPGCSFSIFILA 180
 Db 121 KVIDLTTCGFAIDLGYIPVVGAPLGGAALAHGVRAIEDGINYATGNLPGCSFSIFILA 180
 QY 181 LLSCLTTPASA 191
 Db 181 LLSCLTTPASA 191

RESULT 21
 US-08-836-075A-42 Application US/08836075A
 Patent No. 6180768
 GENERAL INFORMATION:
 APPLICANT: MAERTENS, GEERT
 ADDRESS: ARNOLD, WHITE & DURKEE
 STREET: P.O. BOX 4433
 CITY: HOUSTON
 STATE: TEXAS
 COUNTRY: USA
 ZIP: 77210-4433
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Microsoft Word 6.0 / ASCII text output
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/836,075A
 FILING DATE: 21 Apr 1997
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: PCT/EP95/04155
 FILING DATE: 23 Oct 1995
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: EP 94870166.9
 FILING DATE: 21 Oct 1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: EP 95870076.7
 FILING DATE: 28 Jun 1995
 ATTORNEY/AGENT INFORMATION:
 NAME: KAMMERER, PATRICIA A.
 REGISTRATION NUMBER: 29,775
 INFORMATION FOR SEQ ID NO: 42:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 319 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 US-08-836-075A-42

Query Match 93.5%; Score 977; DB 3; Length 319;
 Best Local Similarity 93.7%; Pred. No. 8, 8e-50; Indels 0; Gaps 0;
 Matches 179; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

QY 1 MSTLPKPKRKTKRNTNRPTDVKPGGGQIVGGYLRRRGPRGRKTRKRSERSQRG 60
 Db 1 MSTLPKPKRKTKRNTNRPTDVKPGGGQIVGGYLRRRGPRGRKTRKRSERSQRG 60
 QY 61 RROPIPKARQPOQRHWAQPGYWPPLYGNBGGWAGWILSPRSRPRPHGPNDRRRSNLG 120
 Db 61 RROPIPKRQPTORSWQPGYWPPLYGNBGGWAGWILSPRSRPRPHGPNDRRRSNLG 120
 QY 121 KVIDLTTCGFAIDLGYIPVVGAPLGGAALAHGVRAIEDGINYATGNLPGCSFSIFILA 180
 Db 121 KVIDLTTCXLAIDLGYIPVVLGGPLGGVAALAHGVRAIEDGVNYATGNLPGCSFSIFILA 180
 QY 181 LLSCLTTPASA 191
 Db 181 LLSCLTTPASA 191

RESULT 22
 US-08-635-886C-229
 Sequence 229, Application US/08635886C
 Patent No. 6155114
 GENERAL INFORMATION:
 APPLICANT: LEROUX-ROELS, Geert
 APPLICANT: DELEYNS, Robert
 APPLICANT: MAERTENS, Geert
 TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C VIRUS
 TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
 TITLE OF INVENTION: PCT/EP94/0555
 CURRENT APPLICATION NUMBER: US/08/635, 886C
 CURRENT FILING DATE: 1994-10-28
 PRIOR APPLICATION NUMBER: EP 93402718.6
 PRIOR FILING DATE: 1993-11-04
 NUMBER OF SEQ ID NOS: 286
 SOFTWARE: Patentin version 3.1
 SEQ ID NO: 229
 LENGTH: 319
 TYPE: PCT
 ORGANISM: hepatitis C virus
 FEATURE: MISC FEATURE
 NAME/KEY: MISC FEATURE
 LOCATION: (128)..(129)
 OTHER INFORMATION: Xaa is any amino acid
 FEATURE: MISC FEATURE
 NAME/KEY: MISC FEATURE
 LOCATION: (310)..(310)
 OTHER INFORMATION: Xaa is any amino acid
 OTHER INFORMATION: Xaa is any amino acid
 NAME/KEY: MISC FEATURE
 LOCATION: (310)..(310)
 OTHER INFORMATION: Xaa is any amino acid
 US-08-635-886C-229

Query Match 93.5%; Score 977; DB 4; Length 319;
 Best Local Similarity 93.7%; Pred. No. 8, 8e-50; Indels 0; Gaps 0;
 Matches 179; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

QY 1 MSTLPKPKRKTKRNTNRPTDVKPGGGQIVGGYLRRRGPRGRKTRKRSERSQRG 60
 Db 1 MSTLPKPKRKTKRNTNRPTDVKPGGGQIVGGYLRRRGPRGRKTRKRSERSQRG 60
 QY 61 RROPIPKARQPOQRHWAQPGYWPPLYGNBGGWAGWILSPRSRPRPHGPNDRRRSNLG 120
 Db 61 RROPIPKRQPTORSWQPGYWPPLYGNBGGWAGWILSPRSRPRPHGPNDRRRSNLG 120
 QY 121 KVIDLTTCGFAIDLGYIPVVGAPLGGAALAHGVRAIEDGINYATGNLPGCSFSIFILA 180
 Db 121 KVIDLTTCXLAIDLGYIPVVLGGPLGGVAALAHGVRAIEDGVNYATGNLPGCSFSIFILA 180
 QY 181 LLSCLTTPASA 191
 Db 181 LLSCLTTPASA 191

RESULT 23
 US-08-974-690C-229
 Sequence 229, Application US/08974690C
 Patent No. 661333
 GENERAL INFORMATION:
 APPLICANT: LEROUX-ROELS, Geert
 APPLICANT: DELEYNS, Robert
 APPLICANT: MAERTENS, Geert
 TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
 TITLE OF INVENTION: VIRUS
 FILE REFERENCE: 2551-94

CURRENT APPLICATION NUMBER: US/08/974,690C
 CURRENT FILING DATE: 1997-11-19
 PRIOR APPLICATION NUMBER: PCT/EP94/03555
 PRIOR FILING DATE: 1994-10-28
 PRIOR APPLICATION NUMBER: EP 93402718.6
 NUMBER OF SEQ ID NOS: 286
 SOFTWARE: Patentin version 3.1
 SEQ ID NO: 229
 LENGTH: 319

TYPE: PRT
 ORGANISM: hepatitis C virus
 FEATURE:
 NAME/KEY: MISC FEATURE
 LOCATION: (128)..(129)
 OTHER INFORMATION: Xaa is any amino acid

FEATURE:
 NAME/KEY: MISC FEATURE
 LOCATION: (308)..(308)
 OTHER INFORMATION: Xaa is any amino acid

FEATURE:
 NAME/KEY: MISC FEATURE
 LOCATION: (310)..(310)
 OTHER INFORMATION: Xaa is any amino acid

US-08-974-690C-229

Query Match
 Best Local Similarity 93.5%; Score 977; DB 4; Length 319;
 Matches 179; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

Qy 1 MSTLPKPKRKTKRNTNRPTDKFPGGQIQVGGVYLPRRGPRGRTRAKTSERSQRG 60
 Db 1 MSTLPKPKRKTKRNTNRPTDKFPGGQIQVGGVYLPRRGPRGRTRAKTSERSQRG 60

Qy 61 RQQPPIKARQPOGRHMAOPGIPWPLYNGEBCGAWGNLSPRSRSPWGPNDRRRSNLG 120
 Db 61 RQQPPIKARQPOGRHMAOPGIPWPLYNGEBCGAWGNLSPRSRSPWGPNDRRRSNLG 120

Qy 121 KVIDLTGCFADIMGYIPVVGAPLGVAALAHGVRAIEDGVINYATGNLPGCSFSIFLLA 180
 Db 121 KVIDLTGCFADIMGYIPVVGAPLGVAALAHGVRAIEDGVINYATGNLPGCSFSIFLLA 180

Qy 181 LISCLTPASA 191
 Db 181 LISCLTPASA 191

RESULT 25
 US-08-974-690C-181
 Sequence 181, Application US/08974690C
 ; Patent No. 6613333

GENERAL INFORMATION:
 APPLICANT: LEROUX-ROELS, Geert
 APPLICANT: DELEYNS, Robert
 APPLICANT: MAERTENS, Geert
 TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPE OF HEPATITIS C
 FILE REFERENCE: 2551-94
 CURRENT APPLICATION NUMBER: US/08/974,690C
 PRIOR FILING DATE: 1997-11-19
 PRIOR APPLICATION NUMBER: PCT/EP94/03555
 PRIOR FILING DATE: 1994-10-28
 NUMBER OF SEQ ID NOS: 286
 SOFTWARE: Patentin version 3.1
 SEQ ID NO: 181
 LENGTH: 450
 TYPE: PRT
 ORGANISM: hepatitis C virus

US-08-974-690C-181

Query Match
 Best Local Similarity 93.5%; Score 977; DB 4; Length 450;
 Matches 176; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

Qy 1 MSTLPKPKRKTKRNTNRPTDKFPGGQIQVGGVYLPRRGPRGRTRAKTSERSQRG 60
 Db 1 MSTLPKPKRKTKRNTNRPTDKFPGGQIQVGGVYLPRRGPRGRTRAKTSERSQRG 60

Qy 61 RQQPPIKARQPOGRHMAOPGIPWPLYNGEBCGAWGNLSPRSRSPWGPNDRRRSNLG 120
 Db 61 RQQPPIKARQPOGRHMAOPGIPWPLYNGEBCGAWGNLSPRSRSPWGPNDRRRSNLG 120

Qy 121 KVIDLTGCFADIMGYIPVVGAPLGVAALAHGVRAIEDGVINYATGNLPGCSFSIFLLA 180
 Db 121 KVIDLTGCFADIMGYIPVVGAPLGVAALAHGVRAIEDGVINYATGNLPGCSFSIFLLA 180

Qy 181 LISCLTPASA 191
 Db 181 LISCLTPASA 191

RESULT 26
 US-08-466-975A-23
 Sequence 23, Application US/08466975A
 ; Patent No. 5910404
 ; GENERAL INFORMATION:
 ; APPLICANT: DELVIS, ROBERT J
 ; APPLICANT: POLLAT, DIRK

US-08-635-886C-181

Query Match
 93.5%; Score 977; DB 4; Length 450;

RESULT 27
US-08-391-671A-23

RESULT 28
 US-08-467-902A-23
 Sequence 23, Application US/08467902A
 Patent No. 6007982
 GENERAL INFORMATION:
 APPLICANT: DELVIS, ROBERT J
 APPLICANT: POLLET, DIRK
 APPLICANT: MAERTENS, GEERT
 APPLICANT: VAN HEUVERSWIJN, HUGO
 TITLE OF INVENTION: SYNTHETIC ANTIGENS FOR THE DETECTION OF
 NUMBER OF SEQUENCES: 23
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: NIXON & VANDERHVE P.C.
 STREET: 1100 NORTH GLEBE ROAD
 CITY: ARLINGTON
 STATE: VA
 COUNTRY: USA
 ZIP: 22201
 COMPUTER READABLE FORM:
 MEDIUM TYPE: FLOPPY disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/467,902A
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US/08/391,671
 FILING DATE:
 APPLICATION NUMBER: US 07/920,286
 FILING DATE: 14-OCT-1992
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: WO PCT/EP91/02409
 FILING DATE: 13-DEC-1991
 PRIOR APPLICATION DATA:
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 7038164000
 TELEX: 703816400
 ATTORNEY/AGENT INFORMATION:
 NAME: SADOFF, B.J.
 REGISTRATION NUMBER: 36,663
 REFERENCE/DOCKET NUMBER: 1487-5
 LENGTH: 2894 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 US-08-467-902A-23

Query Match 93.5%; Score 977; DB 3; Length 2894;
 Best Local Similarity 92.1%; Pred. No. 1, 3e-88; Matches 176; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

QY 1 MSTLKPORKTKRNTNRAPTDVKPGGQIVGGYVLLRRGPRPLGVTRKTSERSOPRG 60
 Db 1 MSTIKPKQRKTKRNTNRAPQDKFPGGGQIVGGYVLLRRGPRPLGVTRKTSERSOPRG 60
 Qy 61 RROPIPKARQPOQHWAQDGPQWPLYGLBGGCAGWGLSPRSRSPRSRPHQPNPDRRSRNG 120
 61 RROPIPKRQREGRWTQAGPQWPMLYGLBGGCAGWGLSPRSRSPRSRPHQPNPDRRSRNG 120
 Qy 121 KVIDTTLTGPADIMGYIPVUGAPLGGAVALARGVRAEDGIVATGHLPGPSFSIIL 180
 121 KVIDTTLTGPADIMGYIPVUGAPLGGAVALARGVRAEDGIVATGHLPGPSFSIIL 180

RESULT 29
 US-09-275-265-23
 Sequence 23, Application US/09275265
 Patent No. 6287761
 GENERAL INFORMATION:
 APPLICANT: DELVIS, ROBERT J
 APPLICANT: POLLET, DIRK
 APPLICANT: MAERTENS, GEERT
 APPLICANT: VAN HEUVERSWIJN, HUGO
 TITLE OF INVENTION: SYNTHETIC ANTIGENS FOR THE DETECTION OF
 NUMBER OF SEQUENCES: 23
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: NIXON & VANDERHVE P.C.
 STREET: 1100 NORTH GLEBE ROAD
 CITY: ARLINGTON
 STATE: VA
 COUNTRY: USA
 ZIP: 22201
 COMPUTER READABLE FORM:
 MEDIUM TYPE: FLOPPY disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/275,265
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US/08/391,671
 FILING DATE: 21-FEB-1995
 APPLICATION NUMBER: US 07/920,286
 FILING DATE: 14-OCT-1992
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: WO PCT/EP91/02409
 FILING DATE: 13-DEC-1991
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: EP 90124241.2
 FILING DATE: 14-DEC-1990
 ATTORNEY/AGENT INFORMATION:
 NAME: SADOFF, B.J.
 REGISTRATION NUMBER: 36,663
 REFERENCE/DOCKET NUMBER: 1487-5
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 7038164000
 TELEX: 703816400
 ATTORNEY/AGENT INFORMATION:
 NAME: SADOFF, B.J.
 REGISTRATION NUMBER: 36,663
 REFERENCE/DOCKET NUMBER: 1487-5
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 7038164000
 TELEX: 703816400
 INFORMATION FOR SEQ ID NO: 23:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2894 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 US-09-275-265-23

Query Match 93.5%; Score 977; DB 3; Length 2894;
 Best Local Similarity 92.1%; Pred. No. 1, 3e-88; Matches 176; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

Qy 1 MSTLKPORKTKRNTNRAPTDVKPGGQIVGGYVLLRRGPRPLGVTRKTSERSOPRG 60
 Db 1 MSTIKPKQRKTKRNTNRAPQDKFPGGGQIVGGYVLLRRGPRPLGVTRKTSERSOPRG 60
 Qy 61 RROPIPKRQREGRWTQAGPQWPMLYGLBGGCAGWGLSPRSRSPRSRPHQPNPDRRSRNG 120
 61 RROPIPKRQREGRWTQAGPQWPMLYGLBGGCAGWGLSPRSRSPRSRPHQPNPDRRSRNG 120
 Qy 121 KVIDTTLTGPADIMGYIPVUGAPLGGAVALARGVRAEDGIVATGHLPGPSFSIIL 180
 121 KVIDTTLTGPADIMGYIPVUGAPLGGAVALARGVRAEDGIVATGHLPGPSFSIIL 180

RESULT 30
US-09-941-611-23
Sequence 23, Application US/09941611
; Patent No 6576417
; GENERAL INFORMATION:
; APPLICANT: DAILEY, ROBERT J
; MAERTENS, GERT
; VAN HEUVERSWN, HUGO
TITLE OF INVENTION: SYNTHETIC ANTIGENS FOR THE DETECTION OF
ANTIBODIES TO HEPATITIS C VIRUS
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHYE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VA
COUNTRY: USA
ZIP: 22201

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/941,611
FILING DATE: 30-Aug-2001
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/391,671
FILING DATE: 1995-02-21
APPLICATION NUMBER: WO PCT/EP91/02409
FILING DATE: 13-DEC-1991
APPLICATION NUMBER: EP 90124241.2
FILING DATE: 14-DEC-1990

ATTORNEY/AGENT INFORMATION:

NAME: SADOF, B.J.

REGISTRATION NUMBER: 35,663
REFERENCE/DOCKET NUMBER: 1487-5

TELECOMMUNICATION INFORMATION:

TELEPHONE: 7038164000
TELEFAX: 7038164100

INFORMATION FOR SEQ ID NO: 23:

SEQUENCE CHARACTERISTICS:

LENGTH: 2894 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear

MOLCULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
SEQUENCE DESCRIPTION: SEQ ID NO: 23:

US-09-941-611-23

Query Match Best Local Similarity 92.1%; Score 977; DB 4; Length 2894;
Matches 176; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

Qy 121 KVIDTLTGFADIMGYIPVVGAPLGGVAALAHGVRAIEDGNYATGNLPGCSFSIFLLA 180
Db 121 KVIDLTGFADIMGYIPVVGAPLGGVAALAHGVRAIEDGNYATGNLPGCSFSIFLLA 180
Qy 181 LISCLTPASA 191
Db 181 LISCLTPASA 191

This Page Blank (uspto)

Copyright (c) 1993 - 2004 CompuGen Ltd.

GeneCore version 5.1.6

Run on:

May 12, 2004, 09:11:55 ; Search time 39 Seconds

(without alignments)

1545.230 Million cell updates/sec

Title: US-09-084-691B-206

Sequence: 1 MSTLPRPKRKRNRNRRP. CFSFSIFLLAISCLTPASA 191

Scoring table: BLOSUM62

Gapext 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315516202 residues

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 90 summaries

Database : SPTRNBL 25:*

1: sp_archeal:*

2: sp_bacteria:*

3: sp_fungi:*

4: sp_human:*

5: sp_invertebrate:*

6: sp_mammal:*

7: sp_mic:*

8: sp_organelle:*

9: sp_phage:*

10: sp_plant:*

11: sp Rodent:*

12: sp_virus:*

13: sp_vertebrate:*

14: sp_unclassified:*

15: sp_rvirus:*

16: sp_bacteriap:*

17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query | Match Length | DB ID | Description |
|------------|-------|-------|--------------|-------|--------------------|
| 1 | 1045 | 100.0 | 191 | 12 | Q68115 hepatitis c |
| 2 | 1035 | 99.0 | 191 | 12 | Q68411 hepatitis c |
| 3 | 1025 | 98.1 | 414 | 12 | P09956 hepatitis c |
| 4 | 1024 | 98.0 | 414 | 12 | Q81329 hepatitis c |
| 5 | 1024 | 98.0 | 414 | 12 | P09958 hepatitis c |
| 6 | 1024 | 98.0 | 3016 | 12 | Q92531 hepatitis c |
| 7 | 1023 | 97.9 | 3018 | 12 | Q39927 hepatitis c |
| 8 | 1020 | 97.6 | 414 | 12 | P89963 hepatitis c |
| 9 | 1016 | 97.2 | 414 | 12 | P89957 hepatitis c |
| 10 | 1016 | 97.2 | 414 | 12 | P89960 hepatitis c |
| 11 | 1014 | 97.0 | 414 | 12 | P89959 hepatitis c |
| 12 | 1011 | 96.7 | 414 | 12 | Q81270 hepatitis c |
| 13 | 1009 | 96.5 | 319 | 12 | Q68703 hepatitis c |
| 14 | 1009 | 96.5 | 414 | 12 | P89962 hepatitis c |
| 15 | 1007 | 96.4 | 191 | 12 | Q93057 hepatitis c |
| 16 | 1007 | 96.4 | 414 | 12 | Q81269 hepatitis c |

90 975 93.3 191 12 Q8v829 hepatitis c Q8v829 hepatitis c

ALIGNMENTS

RESULT 1

ID Q68115 PRELIMINARY; PRT; 191 AA.

AC 068115;

DT 01-NOV-1995 (TREMBrel. 01, Created)

DT 01-NOV-1995 (TREMBrel. 01, Last sequence update)

DT 01-JUN-2003 (TREMBrel. 24, Last annotation update)

DE Genome polyprotein (Fragment).

GN Hepatitis C virus.

OS Hepatitis C virus.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Hepacivirus.

OX NCBI_Taxid=11103;

RN [1]

SEQUENCE FROM N.A.

RC STRAIN=HK2;

RX MEDLINE=9436721; PubMed=805877;

RA Buchk J., Purcell R.H.; Miller R.H.;

RT "Sequence analysis of the core gene of 14 hepatitis C virus genotypes.", Proc. Natl. Acad. Sci. U.S.A. 91:8239-8243(1994). [2]

RP SEQUENCE FROM N.A.

RC STRAIN=HK2;

RX MEDLINE=92279243; PubMed=317578;

RA Buchk J., Purcell R.H., Miller R.H.;

RT "Sequence analysis of the 5' noncoding region of hepatitis C virus.", Proc. Natl. Acad. Sci. U.S.A. 89:4942-4946(1992). [3]

RP SEQUENCE FROM N.A.

RC STRAIN=HK2;

RX MEDLINE=93376778; PubMed=8396266;

RA Buchk J., Purcell R.H., Miller R.H.;

RT "At least 12 genotypes of hepatitis C virus predicted by sequence analysis of the putative E1 gene of isolates collected worldwide.", Proc. Natl. Acad. Sci. U.S.A. 90:8224-8228(1993).

CC -- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA (BY SIMILARITY).

DR EMBL: U33435; RAB40038.1; -.

DR GO: GO:0019028; C:viral capsid; IEA.

DR GO: GO:005138; F:structural molecule activity; IEA.

DR InterPro: IPR00522; HCV core.

DR InterPro: IPR002521; HCV core.

DR Pfam: PF01543; HCV_capsid; 1.

DR Pfam: PF01542; HCV_core; 1.

KW Polyprotein.

FT NON_TER 191 191

SQ SEQUENCE 191 AA; 20725 MW; F292AB64B56DE30A CRC64;

Query Match 99.0%; Score 1035; DB 12; Length 191; Best Local Similarity 99.0%; Pred. No. 1.4e+66; Matches 189; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 MSLPKPKRKTKRNTNRPTDKFPGGGWLLPQQGPRGLGRATKTSERSQPRG 60

DB 1 MSLPKPKRKTKRNTNRPTDKFPGGGWLLPQQGPRGLGRATKTSERSQPRG 60

OY 61 RQQIPKQRQPOGRHWAQPGYPMPLYGNEGCGWAGWLLSPRSRSPWGPNDRRRSRLG 120

DB 61 RQQIPKQRQPOGRHWAQPGYPMPLYGNEGCGWAGWLLSPRSRSPWGPNDRRRSRLG 120

OY 121 KVDIILTCGFADLMGYIPVGAPLGVAALAHGVIAEDGINATGNCFSFLA 180

DB 121 KVDIILTCGFADLMGYIPVGAPLGVAALAHGVIAEDGINATGNCFSFLA 180

FT NON_TER 191 191

SQ SEQUENCE 191 AA; 20725 MW; 7A7994DF2F0909C6 CRC64;

Query Match 100.0%; Score 1045; DB 12; Length 191; Best Local Similarity 100.0%; Pred. No. 1.7e+67; Mismatches 0; Indels 0; Gaps 0;

OY 181 LLSCLTPASA 191

DB 181 LLSCLTPASA 191

RESULT 3

P89956 PRELIMINARY; PRT; 414 AA.

ID P89956

AC P89956;

DT 01-MAY-1997 (TREMBrel. 03, Created)

DT 01-MAY-1997 (TREMBrel. 03, Last sequence update)

DT 01-OCT-2003 (TREMBrel. 25, Last annotation update)

DB Core, env and part of E2/N31 (Genome polyprotein) (Fragment).

OS Hepatitis C virus.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Hepacivirus.

OX NCBI_Taxid=11103;

| | |
|---|---|
| RN [2] | RA |
| RP SEQUENCE FROM N.A. | RA |
| RC STRAIN=VN530; | RA Mayumi M.; |
| RX MEDLINE=95062197; PubMed=7972001; | RT "The entire nucleotide sequences of three hepatitis C virus isolates in genetic groups 7-9 and comparison with those in the other eight genetic groups"; |
| RA Tokita H., Okamoto H., Tsuchida F., Song P., Nakata S., Chosa T., RT "Hepatitis C virus variants from Vietnam are classifiable into the seventh, eighth, and ninth major genetic groups."; | |
| RT Prog. Natl. Acad. Sci. U.S.A. 91:11022-11026(1994). | RL |
| CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: | RL Gen. Virol. 79:1847-1857(1998). |
| CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF CC PROTEIN C AND MENA (BY SIMILARITY). | RL Lipoprotein; EMBL; DB8424; BRA32466.1; -. |
| CC PROTEIN C AND MENA (BY SIMILARITY). | DR PIR; P00894; P00894. |
| CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF CC PROTEIN C AND MENA (BY SIMILARITY). | DR HSSP; P27959; IAV. |
| CC PROTEIN C AND MENA (BY SIMILARITY). | DR MEROPS; S220.001; -. |
| DR EMBL; D88471; BRA1360.1; -. | DR GO; GO:0016021; C:integral to membrane; IEA. |
| DR InterPro; IPR002521; HCV_core. | DR GO; GO:0019028; C:viral capsid; IEA. |
| DR GO; GO:0019031; C:viral envelope; IEA. | DR GO; GO:0019031; C:viral capsid; IEA. |
| DR InterPro; IPR002531; HCV_env. | DR GO; GO:001928; C:viral envelope; IEA. |
| DR InterPro; IPR002531; HCV_Ns1. | DR GO; GO:001928; C:viral capsid; IEA. |
| DR PFam; PF01543; HCV_capsid; 1. | DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA. |
| DR PFam; PF01542; HCV_core; 1. | DR GO; GO:0008236; F:serine-type peptidase activity; IEA. |
| DR PFam; PF01539; HCV_env; 1. | DR GO; GO:0005524; F:ATP binding; IEA. |
| DR PFam; PF01560; HCV_Ns1; 1. | DR GO; GO:0008026; F:ATP dependent helicase activity; IEA. |
| DR Coat protein; Envelope protein; Glycoprotein; Nonstructural protein; KW Polyprotein; Transmembrane. | DR GO; GO:0017678; F:hydrodase activity; IEA. |
| DR NON_TER 414 | DR GO; GO:0003723; F:RNA binding; IEA. |
| FT SEQNER 414 AA; 4973 MW; F3F3CP154372F4FF CRC64; | DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA. |
| Query Match 98.0%; Score 1024; DB 12; Length 414; | DR GO; GO:0005198; F:structural molecule activity; IEA. |
| BEST Local Similarity 97.4%; Pred. No. 3.3e-85; Matches 186; Conservative 2; Mismatches 3; Indels 0; Gaps 0; | DR GO; GO:0016740; F:transf erase activity; IEA. |
| QY 1 MSLPLKQQRKTGKNTNRPTDVKPGGGQIVGGVYLPRRGPRIGRATRKTSERSQRSQRG 60 | DR GO; GO:0006350; P:transcription; IEA. |
| Db 1 MSTLPKQQRKTGKNTNRPTDVKPGGGQIVGGVYLPRRGPRIGRATRKTSERSQRSQRG 60 | DR GO; GO:0019079; P:viral genome replication; IEA. |
| Qy 61 RRQPPIKARQOSGRHWAQPGWPLWYNEGCCWAGMILSPRSRPHGPNNPQQRRRNLG 120 | DR GO; GO:0019087; P:viral transformation; IEA. |
| Db 61 RRQPPIKARQOSGRHWAQPGWPLWYNEGCCWAGMILSPRSRPHGPNNPQQRRRNLG 120 | DR GO; GO:0019091; P:viral genome replication; IEA. |
| Qy 121 KVIDLTTCGPAIDLQGYIPVNGAIPGLGGVAAALHGVRALIEDGNYATGNLPGGSFSIPLLA 180 | DR GO; GO:0019093; CYS_Ser_trypsin. |
| Db 121 KVIDLTTCGPAIDLQGYIPVGLAPLGVVAALAHGVVRVIEDGNYATGNLPGGSFSIPLLA 180 | DR InterPro; IPR002510; DEAD. |
| Qy 181 LMSCLTPASA 191 | DR InterPro; IPR002522; HCV_capsid. |
| Db 181 LMSCLTPASA 191 | DR InterPro; IPR002521; HCV_core. |
| RESULT 6 | DR InterPro; IPR002531; HCV_env. |
| 092531 ID PRELIMINARY; PRT; 3016 AA. | DR InterPro; IPR002531; HCV_Ns1. |
| AC 092531; DT 01-NOV-1998 (TREMBLrel. 08, Created) | DR InterPro; IPR002518; HCV_Ns2. |
| DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update) | DR InterPro; IPR002519; HCV_Ns4. |
| DE Genome polyprotein. | DR InterPro; IPR002531; HCV_RarP. |
| OS Hepatitis C virus. | DR InterPro; IPR002531; HCV_Ns5. |
| OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; RC Hepacivirus. | DR InterPro; IPR002531; HCV_Ns5. |
| OC Hepacivirus. | DR InterPro; IPR002531; HCV_Ns5. |
| OX NCBI_TaxID=1103; RN [1] | DR InterPro; IPR002531; HCV_Ns5. |
| RP SEQUENCE FROM N.A. | DR InterPro; IPR002531; HCV_Ns5. |
| RC STRAIN=VN405; RA Okamoto H.; DR SMART; SM0047; DEXO; 1. | DR InterPro; IPR002531; HCV_Ns5. |
| RA Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases. | DR InterPro; IPR002531; HCV_Ns5. |
| RP SEQUENCE FROM N.A. | DR InterPro; IPR002531; HCV_Ns5. |
| RC STRAIN=VN405; RX MEDLINE=98378034; PubMed=9714232; | DR InterPro; IPR002531; HCV_Ns5. |
| Query Match 98.0%; Score 1024; DB 12; Length 3016; | DR InterPro; IPR002531; HCV_Ns5. |
| BEST Local Similarity 98.4%; Pred. No. 3.1e-84; Matches 188; Conservative 1; Mismatches 2; Indels 0; Gaps 0; | DR InterPro; IPR002531; HCV_Ns5. |
| QY 1 MSLPLKQQRKTGKNTNRPTDVKPGGGQIVGGVYLPRRGPRIGRATRKTSERSQRSQRG 60 | DR InterPro; IPR002531; HCV_Ns5. |
| Db 61 RRQPPIKARQOSGRHWAQPGWPLWYNEGCCWAGMILSPRSRPHGPNNPQQRRRNLG 120 | DR InterPro; IPR002531; HCV_Ns5. |
| Qy 61 RRQPPIKARQOSGRHWAQPGWPLWYNEGCCWAGMILSPRSRPHGPNNPQQRRRNLG 120 | DR InterPro; IPR002531; HCV_Ns5. |

| | | | | |
|-----------|---|---|--------------|--|
| QY | 181 | LISCLTPASA | 191 | OC Hepacivirus. |
| OX | | | | NCBI_TaxID=1103; |
| RN | | | | [1] SEQUENCE FROM N.A.. |
| RP | | | | SEQUENCE FROM N.A.. |
| RC | | | | STRAIN=VN569; |
| RA | | | | Okamoto H.'s |
| RL | | | | Submitted (AUG-1996) to the EMBL/GenBank/DDJB databases. |
| RN | | | | [2] |
| ID | 068703 | PRELIMINARY; | PRT; 319 AA. | RN |
| AC | Q68703; | | | SEQUENCE FROM N.A.. |
| DT | 01-NOV-1996 | (TREMBLrel. 01, Created) | | RN |
| DT | 01-NOV-1996 | (TREMBLrel. 01, Last sequence update) | | STRAIN=VN569; |
| DB | Core protein (fragment). | | | RN |
| RT | 01-JUN-2003 | (TREMBLrel. 24, Last annotation update) | | RA |
| OS | Hepatitis C virus type 6a. | | | Iizuka H., Misiro S., Miyakawa Y., Mayumi M., Nakata S., Chosa T., Tokita H., Okamoto H., Truda F., Song P., |
| OC | Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; | | | "Hepatitis C virus variants from Vietnam are classifiable into the seven, eighth, and ninth major genetic groups"; Proc. Natl. Acad. Sci. U.S.A. 91:1102-1106(1994). |
| OC | Hepacivirus. | | | -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: CC-OC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF CC-CC PROTEIN C AND MRNA (BY SIMILARITY). |
| RN | | | | EMBL; D88475; BRA13624.1; -. |
| RA | Suyver L., Wyseur A., van Arnhem W., Lunel F., Laurent-Puig P., | | | DR GO; GO:0016021; C:integral to membrane; IEA. |
| RA | Pawlowsky J.-M., Kleter B., Babbit L., Nkengabong J., van Doorn L.-J., Maertens G.; | | | DR GO; GO:0019028; C:viral capsid; IEA. |
| RT | "Hepatitis C virus genotyping by means of 5'-UR/core line probe assays and molecular analysis of untypeable samples."; | | | DR GO; GO:0005198; F:structural molecule activity; IEA. |
| RL | Virus Res. 38:137-157(1995). | | | DR GO; GO:0019031; C:viral envelope; IEA. |
| DR | EMBL; L38339; AAC42193.1; -. | | | DR GO; GO:0005198; F:structural molecule activity; IEA. |
| DR | GO; GO:0019028; C:viral capsid; IEA. | | | DR InterPro; IPR002522; HCV_capsid. |
| DR | GO; GO:0019031; C:viral envelope; IEA. | | | DR InterPro; IPR002521; HCV core. |
| DR | PFam; PF01543; HCV_capsid; 1. | | | DR InterPro; IPR002519; HCV env. |
| DR | PFam; PF01542; HCV_core; 1. | | | DR InterPro; IPR002531; HCV_NS1. |
| DR | PFam; PF01539; HCV_env; 1. | | | DR Pfam; PF01543; HCV capsid; 1. |
| DR | PFam; PF01539; HCV_env; 1. | | | DR Pfam; PF01539; HCV_env; 1. |
| DR | PFam; PF01560; HCV_NS1. | | | DR Pfam; PF01560; HCV_NS1. |
| KW | Coat protein; Envelope protein; Glycoprotein; Nonstructural protein; Polyprotein; Transmembrane. | | | DR Pfam; PF01542; HCV_core; 1. |
| FT | NON_TER 319 | | | DR Pfam; PF01539; HCV_env; 1. |
| FT | NON_TER 319 AA; 34545 MW; DC1D0BA52ED64A8F CRC64; | | | DR Pfam; PF01539; HCV_env; 1. |
| FT | Query Match 96.6%; Score 1009; DB 12; Length 319; Best Local Similarity 96.3%; Pred. No. 5. 8e-84; Matches 184; Conservative 2; Mismatches 5; Indels 0; Gaps 0; | | | DR Pfam; PF01539; HCV_env; 1. |
| QY | 1 | MSTLPKQORKTKRNTTRPMDVKPGQQIVGVVYLPRRGRLGVATRKTSERSOPRG | 60 | DR Pfam; PF01539; HCV_env; 1. |
| QY | 1 | MSTLPKQORKTKRNTTRPMDVKPGQQIVGVVYLPRRGRLGVATRKTSERSOPRG | 60 | DR Pfam; PF01539; HCV_env; 1. |
| QY | 61 | RQQPPIKARQOGRHWAQPGYWPPLYGNEGCGCAGWLLSPRSRPRNGPNDRRRSRLG | 120 | DR Pfam; PF01539; HCV_env; 1. |
| Db | 61 | RQQPPIKARQOGRHWAQPGYWPPLYGNEGCGCAGWLLSPRSRPRNGPNDRRRSRLG | 120 | DR Pfam; PF01539; HCV_env; 1. |
| QY | 121 | KVIDLTCGGFADLMGYIPVVGAPLGGVAAALAHGVRAIEDGNYATCNLPGCSFSFLA | 180 | DR Pfam; PF01539; HCV_env; 1. |
| Db | 121 | KVIDLTCGGFADLMGYIPVVGAPLGGVAAALAHGVRAIEDGNYATCNLPGCSFSFLA | 180 | DR Pfam; PF01539; HCV_env; 1. |
| QY | 181 | LISCLTPASA 191 | | DR Pfam; PF01539; HCV_env; 1. |
| Db | 181 | LISCLTPASA 191 | | DR Pfam; PF01539; HCV_env; 1. |
| RESULT 14 | | | | |
| P89962 | | | | |
| ID | P89962 | PRELIMINARY; | PRT; 414 AA. | RESULT 15 |
| AC | P89962; | | | 093057 PRELIMINARY; PRT; 191 AA. |
| DT | 01-MAY-1997 | (TREMBLrel. 03, Created) | | ID 093057; Q68747; PRELIMINARY; PRT; 191 AA. |
| DT | 01-MAY-1997 | (TREMBLrel. 03, Last sequence update) | | AC 093057; Q68747; PRELIMINARY; PRT; 191 AA. |
| DT | 01-JUN-2003 | (TREMBLrel. 24, Last annotation update) | | DT 01-NOV-1998 (TREMBLrel. 08, Created) |
| DE | Core, env and part of E2/NS1 (Genome polyprotein) (Fragment). | | | DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update) |
| OS | Hepatitis C virus. | | | DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update) |
| OC | Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; | | | OS Hepatitis C virus. |
| OC | Hepacivirus. | | | OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; |
| OC | Reoviridae. | | | OC Hepacivirus. |
| OC | Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; | | | OC NCBI_TaxID=1103; [1] SEQUENCE FROM N.A.. |

RC STRAINS-D97/93;
 RX MEDLINE=97060095; PubMed=8904416;
 RA doi: H. Apichartpiyakul C., Ohba K., Mizokami M., Rotta H.; Thailand,
 RT "hepatitis C virus (HCV) subtype prevalence in Chiang Mai, Thailand,
 RT and identification of novel subtypes of HCV major types 6.";
 RL J. Clin. Microbiol. 34:569-574(1996);
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 PROTEIN C AND MRNA (BY SIMILARITY).
 DR EMBL: D63946; BAA09976.1; -.
 GO: GO:0015028; C:viral capsid; IEA.
 DR GO: GO:0005198; F:structural molecule activity; IEA.
 DR InterPro; IPR002522; HCV capsid.
 DR InterPro; IPR002521; HCV core.
 DR InterPro; IPR002519; HCV env.
 DR InterPro; IPR002531; HCV_NSL.
 DR Pfam; PF01543; HCV_capsid; 1.
 DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV_env; 1.
 DR Pfam; PF01560; HCV_NSL; 1.
 KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
 Polyprotein; Transmembrane.
 FT NON_TER 191 191
 SF SEQUENCE 191 AA; 20727 MW; 411C1C851ADA2DB5 CRC64;
 QY Best Local Similarity 96.4%; Score 1007; DB 12; Length 191;
 Matches 182; Conservative 5; Mismatches 4; Indels 0; Gaps 0;
 QY 1 MSTLPKPORKTAKNTNRPPTDKPGGQIVGAVYLAPRGRPLGRVATRKTSERSOPRG 60
 Db 1 MSTLPKPORKTAKNTNRPPTDKPGGQIVGAVYLAPRGRPLGRVATRKTSERSOPRG 60
 QY 61 RROPIPKARQPOCRHWAQPGYWPMLYNGEGCWAGWLLSPRSRPNWGPNDRRSNLG 120
 Db 61 RROPIPKARQPOCRHWAQPGYWPMLYNGEGCWAGWLLSPRSRPNWGPNDRRSNLG 120
 QY 121 KVDTLTCGFADLMGYIPVVGAPLGVAALAHGVRAIEDGINATGNLPGCSFSIFILA 180
 Db 121 KVDTLTCGFADLMGYIPVVGAPLGVAALAHGVRAIEDGINATGNLPGCSFSIFILA 180
 QY 181 LFSCLTTPASA 191
 Db 181 LFSCLTTPASA 191
 RESULT 16
 Q81269 ID 081269 PRELIMINARY; PRT; 414 AA.
 AC 081269;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-NOV-1996 (TREMBlrel. 01, Last annotation update)
 DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update), (Fragment).
 DR 01-JUN-2003 (TREMBlrel. 24, Last annotation update), (Fragment).
 OS Core, env and part of E2/NS1 (Genome polyprotein) (Fragment).
 OC Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Replicaviridae;
 OC NCBI_TAXID-1103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Th271;
 RX MEDLINE=96005057; PubMed=7561773;
 RA Tokita H., Okamoto H., Luengrojanakul P., Vareesangthip K.,
 RA Chainnati T., Iizuka H., Tsuda F., Miyakawa Y., Mayumi M.;
 RT "Hepatitis C virus variants from Thailand classifiable into five novel
 genotypes in the sixth (6b), seventh (7c,7d) and ninth (9b, 9c) major
 genetic groups." J. Gen. Virol. 76:2329-2335(1995).
 RT genotypes in the sixth (6b), seventh (7c,7d) and ninth (9b, 9c) major
 RT genotypes in the sixth (6b), seventh (7c,7d) and ninth (9b, 9c) major
 RT genotypes in the sixth (6b), seventh (7c,7d) and ninth (9b, 9c) major
 RT genotypes in the sixth (6b), seventh (7c,7d) and ninth (9b, 9c) major
 RT genotypes in the sixth (6b), seventh (7c,7d) and ninth (9b, 9c) major
 RT genotypes in the sixth (6b), seventh (7c,7d) and ninth (9b, 9c) major
 RT genotypes in the sixth (6b), seventh (7c,7d) and ninth (9b, 9c) major
 RT genotypes in the sixth (6b), seventh (7c,7d) and ninth (9b, 9c) major
 RT genotypes in the sixth (6b), seventh (7c,7d) and ninth (9b, 9c) major
 RT genotypes in the sixth (6b), seventh (7c,7d) and ninth (9b, 9c) major
 DR EMBL: D37843; BAA07089.1; -.
 DR InterPro; IPR002521; HCV core.
 DR InterPro; IPR002519; HCV env.
 DR InterPro; IPR002531; HCV_NSL.
 DR Pfam; PF01543; HCV_capsid; 1.
 DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV_env; 1.
 DR Pfam; PF01560; HCV_NSL; 1.
 SF SEQUENCE 414 AA; 45072 MW; 6C9E1AFAE4642241D CRC64;
 QY Query Match 96.4%; Score 1007; DB 12; Length 414;
 Best Local Similarity 96.9%; Pred. No. 1; 2e-83;
 Matches 185; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
 QY 1 MSTLPKPORKTAKNTNRPPTDKPGGQIVGAVYLAPRGRPLGRVATRKTSERSOPRG 60
 DR 1 MSTLPKPORKTAKNTNRPPTDKPGGQIVGAVYLAPRGRPLGRVATRKTSERSOPRG 60
 QY 61 RROPIPKARQPOCRHWAQPGYWPMLYNGEGCWAGWLLSPRSRPNWGPNDRRSNLG 120
 DR 61 RROPIPKARQPOCRHWAQPGYWPMLYNGEGCWAGWLLSPRSRPNWGPNDRRSNLG 120
 QY 121 KVDTLTCGFADLMGYIPVVGAPLGVAALAHGVRAIEDGINATGNLPGCSFSIFILA 180
 DR 121 KVDTLTCGFADLMGYIPVVGAPLGVAALAHGVRAIEDGINATGNLPGCSFSIFILA 180
 QY 181 LFSCLTTPASA 191
 DR 181 LFSCLTTPASA 191
 RESULT 17
 Q81268 ID 081268 PRELIMINARY; PRT; 414 AA.
 AC 081268;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update), (Fragment).
 DE Core, env and part of E2/NS1 (Genome polyprotein) (Fragment).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Th846;
 RX MEDLINE=96005057; PubMed=7561773;
 RA Tokita H., Okamoto H., Luengrojanakul P., Vareesangthip K.,
 RA Chainnati T., Iizuka H., Tsuda F., Miyakawa Y., Mayumi M.;
 RT "Hepatitis C virus variants from Thailand classifiable into five novel
 genotypes in the sixth (6b), seventh (7c,7d) and ninth (9b, 9c) major
 RT genotypes in the sixth (6b), seventh (7c,7d) and ninth (9b, 9c) major
 RT genotypes in the sixth (6b), seventh (7c,7d) and ninth (9b, 9c) major
 RT genotypes in the sixth (6b), seventh (7c,7d) and ninth (9b, 9c) major
 RT genotypes in the sixth (6b), seventh (7c,7d) and ninth (9b, 9c) major
 RT genotypes in the sixth (6b), seventh (7c,7d) and ninth (9b, 9c) major
 RT genotypes in the sixth (6b), seventh (7c,7d) and ninth (9b, 9c) major
 RT genotypes in the sixth (6b), seventh (7c,7d) and ninth (9b, 9c) major
 RT genotypes in the sixth (6b), seventh (7c,7d) and ninth (9b, 9c) major
 RT genotypes in the sixth (6b), seventh (7c,7d) and ninth (9b, 9c) major
 RT genotypes in the sixth (6b), seventh (7c,7d) and ninth (9b, 9c) major
 RT genotypes in the sixth (6b), seventh (7c,7d) and ninth (9b, 9c) major
 DR EMBL: D37843; BAA07089.1; -.
 DR InterPro; IPR002521; C:integral to membrane; IEA.
 DR GO: GO:0015028; C:viral capsid; IEA.
 DR GO: GO:0019031; C:viral envelope; IEA.
 DR GO: GO:005198; F:structural molecule activity; IEA.
 DR InterPro; IPR002522; HCV capsid.
 DR InterPro; IPR002521; HCV core.
 DR InterPro; IPR002519; HCV env.
 DR InterPro; IPR002531; HCV_NSL.
 DR Pfam; PF01543; HCV_capsid; 1.
 DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV_env; 1.
 DR Pfam; PF01560; HCV_NSL; 1.

| | | | |
|------------------|--|----|---|
| KW | Coat protein; Envelope protein; Glycoprotein; Nonstructural protein; | Qy | 1 MSLPLPKPQRKTKRNTNRRPTDVKFPQGGQIVGGVILLPRRGRLGRTRKTISERSOPRG |
| KW | Polyprotein; Transmembrane. | Db | 1 MSLPLPKPQRKTKRNTNRRPTDVKFPQGGQIVGGVILLPRRGRLGRTRKTISERSOPRG |
| FT | NON_TER 414 AA; 414 MW; 212740491RDA0B1 CRC64; | Qy | 61 RQQPPIKARQPGHWAQPGCPWPPLYNEGGWAGWILSPRSRPHWGPNPDRRSRLG 120 |
| SQ | SEQUENCE 414 AA; 44836 MW; | Db | 61 RQQPPIKARQPGHWAQPGCPWPPLYNEGGWAGWILSPRSRPHWGPNPDRRSRLG 60 |
| Query Match | Best Local Similarity 96.0%; Score 1003; DB 12; Length 414; Matches 185; Conservative 1; Mismatches 5; Indels 0; Gaps 0; | Qy | 1 MSLPLPKPQRKTKRNTNRRPTDVKFPQGGQIVGGVILLPRRGRLGRTRKTISERSOPRG 60 |
| Qy | 1 MSLPLPKPQRKTKRNTNRRPTDVKFPQGGQIVGGVILLPRRGRLGRTRKTISERSOPRG 60 | Db | 1 MSLPLPKPQRKTKRNTNRRPTDVKFPQGGQIVGGVILLPRRGRLGRTRKTISERSOPRG 60 |
| Db | 1 MSLPLPKPQRKTKRNTNRRPTDVKFPQGGQIVGGVILLPRRGRLGRTRKTISERSOPRG 60 | Qy | 61 RQQPPIKARQPGHWAQPGCPWPPLYNEGGWAGWILSPRSRPHWGPNPDRRSRLG 120 |
| Qy | 61 RQQPPIKARQPGHWAQPGCPWPPLYNEGGWAGWILSPRSRPHWGPNPDRRSRLG 120 | Db | 61 RQQPPIKARQPGHWAQPGCPWPPLYNEGGWAGWILSPRSRPHWGPNPDRRSRLG 120 |
| Db | 61 RQQPPIKARQPGHWAQPGCPWPPLYNEGGWAGWILSPRSRPHWGPNPDRRSRLG 120 | Qy | 121 KVIDTUTCGFAIDLGYIPVVGAPLGLGVAAALAHGVRAIEDGINYATCNLPGCSFSIPLLA 180 |
| Qy | 121 KVIDTUTCGFAIDLGYIPVVGAPLGLGVAAALAHGVRAIEDGINYATCNLPGCSFSIPLLA 180 | Db | 121 KVIDTUTCGFAIDLGYIPVVGAPLGLGVAAALAHGVRAIEDGINYATCNLPGCSFSIPLLA 180 |
| Db | 121 KVIDTUTCGFAIDLGYIPVVGAPLGLGVAAALAHGVRAIEDGINYATCNLPGCSFSIPLLA 180 | Qy | 181 LLSCLTPASA 191 |
| Db | 181 LLSCLTPASA 191 | Db | 181 LLSCLTPASA 191 |
| RESULT 18 | | | |
| ID | 081314 PRELIMINARY; PRT; 414 AA. | Qy | 1 MSLPLPKPQRKTKRNTNRRPTDVKFPQGGQIVGGVILLPRRGRLGRTRKTISERSOPRG 60 |
| AC | 081314: | Db | 1 MSLPLPKPQRKTKRNTNRRPTDVKFPQGGQIVGGVILLPRRGRLGRTRKTISERSOPRG 60 |
| DT | 01-NOV-1995 (TREMBLrel. 01, Last sequence update) | Qy | 61 RQQPPIKARQPGHWAQPGCPWPPLYNEGGWAGWILSPRSRPHWGPNPDRRSRLG 120 |
| DT | 01-JUN-1995 (TREMBLrel. 24, Last annotation update) | Db | 61 RQQPPIKARQPGHWAQPGCPWPPLYNEGGWAGWILSPRSRPHWGPNPDRRSRLG 120 |
| DE | Core env and part of E2/NS1 (Genome polyprotein) (Fragment). | Qy | 121 KVIDTUTCGFAIDLGYIPVVGAPLGLGVAAALAHGVRAIEDGINYATCNLPGCSFSIPLLA 180 |
| OS | Hepatitis C virus. | Db | 121 KVIDTUTCGFAIDLGYIPVVGAPLGLGVAAALAHGVRAIEDGINYATCNLPGCSFSIPLLA 180 |
| OC | Hepacivirus. | Qy | 181 LLSCLTPASA 191 |
| OX | NCBI_TaxId:1103; [1] | Db | 181 LLSCLTPASA 191 |
| RP | SEQUENCE FROM N.A. | Qy | 181 LLSCLTPASA 191 |
| RC | STRAIN=Th571; | Db | 181 LLSCLTPASA 191 |
| RX | MEDLINE=9605057; PubMed=7561773; [1] | Qy | 181 LLSCLTPASA 191 |
| RN | 081314: | Db | 181 LLSCLTPASA 191 |
| RN | 01-JUN-2003 (TREMBLrel. 24, Last annotation update) | Qy | 181 LLSCLTPASA 191 |
| RN | 01-AUG-1998 (TREMBLrel. 07, Last sequence update) | Db | 181 LLSCLTPASA 191 |
| RC | SEQUENCE FROM N.A. | Qy | 181 LLSCLTPASA 191 |
| RC | STRAIN=B4/92; | Db | 181 LLSCLTPASA 191 |
| RX | MEDLINE=97060095; PubMed=8904116; [1] | Qy | 181 LLSCLTPASA 191 |
| RA | Doi H.; Apichartpiyakul C., Ohba K., Mizokami M., Hotta H.; Thailand, Hepatitis C virus subtype prevalence in Chiang Mai, Thailand, and identification of novel subtypes of HCV major types 6.;" | Db | 181 LLSCLTPASA 191 |
| RA | J. Clin. Microbiol. 34:569-574(1996). | Qy | 181 LLSCLTPASA 191 |
| RT | "HEPATITIS C VIRUS (HCV) SUBTYPE PREVALENCE IN CHIANG MAI, THAILAND AND IDENTIFICATION OF NOVEL SUBTYPES OF HCV MAJOR TYPES 6." | Db | 181 LLSCLTPASA 191 |
| RT | "SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA (BY SIMILARITY)." | Qy | 181 LLSCLTPASA 191 |
| RT | EMLBL: D63943; BRA09974.1; - | Db | 181 LLSCLTPASA 191 |
| RT | GO; GO:0019028; C:viral capsid; IEA. | Qy | 181 LLSCLTPASA 191 |
| RT | GO; GO:0005198; F:structural molecule activity; IEA. | Db | 181 LLSCLTPASA 191 |
| RT | InterPro: IPR002522; HCV capsid. | Qy | 181 LLSCLTPASA 191 |
| RT | InterPro: IPR002521; HCV core. | Db | 181 LLSCLTPASA 191 |
| RT | PFAM: PF01543; HCV capsid; 1. | Qy | 181 LLSCLTPASA 191 |
| DR | EMBL: D38078; BAB0721.1; - | DR | 181 LLSCLTPASA 191 |
| DR | GO; GO:0019028; C:viral capsid; IEA. | DR | 181 LLSCLTPASA 191 |
| DR | GO; GO:0015031; C:integral to membrane; IEA. | DR | 181 LLSCLTPASA 191 |
| CC | PROTEIN C AND MRNA (BY SIMILARITY). | DR | 181 LLSCLTPASA 191 |
| CC | PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF POLYPROTEIN. | DR | 181 LLSCLTPASA 191 |
| CC | LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF POLYPROTEIN. | DR | 181 LLSCLTPASA 191 |
| CC | PROTEIN C AND MRNA (BY SIMILARITY). | DR | 181 LLSCLTPASA 191 |
| CC | PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF POLYPROTEIN. | DR | 181 LLSCLTPASA 191 |
| Query Match | Best Local Similarity 95.8%; Score 101; DB 12; Length 191; Matches 183; Conservative 3; Mismatches 5; Indels 0; Gaps 0; | Qy | 1 MSLPLPKPQRKTKRNTNRRPTDVKFPQGGQIVGGVILLPRRGRLGRTRKTISERSOPRG 60 |
| Qy | 1 MSLPLPKPQRKTKRNTNRRPTDVKFPQGGQIVGGVILLPRRGRLGRTRKTISERSOPRG 60 | Db | 1 MSLPLPKPQRKTKRNTNRRPTDVKFPQGGQIVGGVILLPRRGRLGRTRKTISERSOPRG 60 |
| Db | 1 MSLPLPKPQRKTKRNTNRRPTDVKFPQGGQIVGGVILLPRRGRLGRTRKTISERSOPRG 60 | Qy | 61 RQQPPIKARQPGHWAQPGCPWPPLYNEGGWAGWILSPRSRPHWGPNPDRRSRLG 120 |
| Qy | 61 RQQPPIKARQPGHWAQPGCPWPPLYNEGGWAGWILSPRSRPHWGPNPDRRSRLG 120 | Db | 61 RQQPPIKARQPGHWAQPGCPWPPLYNEGGWAGWILSPRSRPHWGPNPDRRSRLG 120 |
| Db | 61 RQQPPIKARQPGHWAQPGCPWPPLYNEGGWAGWILSPRSRPHWGPNPDRRSRLG 120 | Qy | 121 KVIDTUTCGFAIDLGYIPVVGAPLGLGVAAALAHGVRAIEDGINYATCNLPGCSFSIPLLA 180 |
| Qy | 121 KVIDTUTCGFAIDLGYIPVVGAPLGLGVAAALAHGVRAIEDGINYATCNLPGCSFSIPLLA 180 | Db | 121 KVIDTUTCGFAIDLGYIPVVGAPLGLGVAAALAHGVRAIEDGINYATCNLPGCSFSIPLLA 180 |
| Db | 121 KVIDTUTCGFAIDLGYIPVVGAPLGLGVAAALAHGVRAIEDGINYATCNLPGCSFSIPLLA 180 | Qy | 181 LLSCLTPASA 191 |
| Db | 181 LLSCLTPASA 191 | Qy | 181 LLSCLTPASA 191 |
| Query Match | Best Local Similarity 96.0%; Score 1003; DB 12; Length 414; Matches 184; Conservative 0; Mismatches 7; Indels 0; Gaps 0; | Qy | 181 LLSCLTPASA 191 |

RESULT 20

ID 081271 PRELIMINARY; PRT; 414 AA.

AC 081271; Core, env and part of E2/NS1 (Genome polyprotein) (Fragment).

DT 01-NOV-1995 (TREMBrel. 01, Last sequence update)

DT 01-JUN-2003 (TREMBrel. 24, Last annotation update)

DE Core, env and part of E2/NS1 (Genome polyprotein) (Fragment).

OS Hepatitis C virus.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OX Hepacivirus.

OX NCBI_TaxID=1103;

RN [1] NCBI_TaxID=1103;

RP SEQUENCE FROM N.A.

RC STRAIN=Th976;

RX MEDLINE=9605057; PubMed=7561773;

RA Tokita H., Okamoto H., Luengrojanakul P., Vareesangthip K., Chainuvati T., Iizuka H., Tsuda F., Miyakawa Y., Mayumi M.; "Hepatitis C virus variants from Thailand classifiable into five novel genotypes in the sixth (6b), seventh (7c/7d) and ninth (9b, 9c) major genetic groups." J. Gen. Virol. 76:2329-2335 (1995).

CC -- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA (BY SIMILARITY).

CC PROTEIN C AND MRNA E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA (BY SIMILARITY).

DR GO: 0016021; C:integral to membrane; IEA.

DR GO: 001931; C:viral envelope; IEA.

DR GO: 0005198; F:structural molecule activity; IEA.

DR EMBL; D37846; BA07032.1; -.

DR GO; GO: 0016021; C:integral to membrane; IEA.

DR GO; GO: 001931; C:viral capsid; IEA.

DR GO; GO: 0005198; F:structural molecule activity; IEA.

DR Interpro; IPR002521; HCV_core.

DR Interpro; IPR002519; HCV_env.

DR Pfam; PF01542; HCV_core; 1.

DR Pfam; PF01560; HCV_NS1; 1.

DR Pfam; PF01539; HCV_NS1; 1.

DR Pfam; PF01542; HCV_env; 1.

DR Interpro; IPR002521; HCV_core.

DR Interpro; IPR002519; HCV_env.

DR Pfam; PF01543; HCV_capsid; 1.

DR Pfam; PF01542; HCV_core; 1.

DR Pfam; PF01539; HCV_env; 1.

DR Pfam; PF01560; HCV_NS1; 1.

KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein; Polyprotein; Transmembrane.

FT NON_TER SEQUENCE 414 AA; 4514 MW; D63EE7CED5B71776 CRC64;

Query Match 95.7%; Score 1000; DB 12; Length 414;

Best Local Similarity 96.3%; Pred. No. 5 2e-83; Matches 184; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Q 1 MSTLPKQORKTRKNTNRPTDKFPGGGQIVGCGVYLPPRGRLGVTRTKTSERSOPRG 60

Db 1 MSTLPKQORKTRKNTNRPTDKFPGGGQIVGCGVYLPPRGRLGVTRTKTSERSOPRG 60

QY 61 RQQPIPKARQPOGRHMAQPGWPWLYNGEGCWAGWGLSPRSRPHGPNDRRSRNLG 120

Db 61 RQQPIPKARQPOGRHMAQPGWPWLYNGEGCWAGWGLSPRSRPHGPNDRRSRNLG 120

QY 121 KVIDTLCGFAIDLGYTIPVVGAPLGGAVALAHLHGVRALEDGNYATGNLPGCSFSIPLLA 180

Db 121 KVIDTLCGFAIDLGYTIPVVGAPLGGAVALAHLHGVRALEDGNYATGNLPGCSFSIPLLA 180

Db 121 KVIDTLCGFAIDLGYTIPVVGAPLGGAVALAHLHGVRALEDGNYATGNLPGCSFSIPLLA 180

QY 181 LISCLTPASA 191

Db 181 LISCLTPASA 191

RESULT 22

P89954 PRELIMINARY; PRT; 414 AA.

ID P89954 Core, env and part of E2/NS1 (Genome polyprotein) (Fragment).

AC P89954; Created)

DT 01-MAY-1997 (TREMBrel. 03, Last sequence update)

DT 01-JUN-2003 (TREMBrel. 24, Last annotation update)

DE Core, env and part of E2/NS1 (Genome polyprotein) (Fragment).

OS Hepatitis C virus.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OX Hepacivirus.

RN [1] NCBI_TaxID=1103;

RP SEQUENCE FROM N.A.

RC STRAIN=VN005;

RX Submitted (AUG-1996) to the EMBL/GenBank/DDBJ databases.

RL RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=VN005;

RX MEDLINE=95062197; PubMed=7972001;

RA Tokita H., Okamoto H., Tsuda F., Song P., Nakata S., Chosa T., Iizuka H., Misiro S., Miyakawa Y., Mayumi M.; "Hepatitis C virus variants from Vietnam are classifiable into the seventh, eighth, and ninth major genetic groups"; Proc. Natl. Acad. Sci. U.S.A. 91:11022-11026 (1994).

CC -I SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA (BY SIMILARITY).

CC DR GO; GO: 0016021; C:integral to membrane; IEA.

DR GO; GO: 001931; C:viral capsid; IEA.

DR GO; GO: 0005198; F:structural molecule activity; IEA.

DR Interpro; IPR002522; HCV_capsid.

DR Interpro; IPR002521; HCV_core.

DR Interpro; IPR002519; HCV_env.

DR Pfam; PF01543; HCV_capsid; 1.

DR Pfam; PF01542; HCV_core; 1.

DR Pfam; PF01539; HCV_env; 1.

DR Pfam; PF01560; HCV_NS1; 1.

KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein; Polyprotein; Transmembrane.

FT NON_TER SEQUENCE 414 AA; 44890 MW; 503E43390BA2663A CRC64;

Query Match 95.6%; Score 999; DB 12; Length 414;

Best Local Similarity 95.3%; Pred. No. 6 4e-83; Matches 182; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 1 MSTLPKQORKTRKNTNRPTDKFPGGGQIVGCGVYLPPRGRLGVTRTKTSERSOPRG 60

Db 1 MSTLPKQORKTRKNTNRPTDKFPGGGQIVGCGVYLPPRGRLGVTRTKTSERSOPRG 60

QY 61 RQQPIPKARQPOGRHMAQPGWPWLYNGEGCWAGWGLSPRSRPHGPNDRRSRNLG 120

Db 61 RQQPIPKARQPOGRHMAQPGWPWLYNGEGCWAGWGLSPRSRPHGPNDRRSRNLG 120

QY 121 KVIDTLCGFAIDLGYTIPVVGAPLGGAVALAHLHGVRALEDGNYATGNLPGCSFSIPLLA 180

Db 121 KVIDTLCGFAIDLGYTIPVVGAPLGGAVALAHLHGVRALEDGNYATGNLPGCSFSIPLLA 180

Db 121 KVIDTLCGFAIDLGYTIPVVGAPLGGAVALAHLHGVRALEDGNYATGNLPGCSFSIPLLA 180

QY 181 LISCLTPASA 191

Db 181 LISCLTPASA 191

OX NCBI_TAXID=1103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=VN04;
 RA Okamoto H.;
 RL Submitted (AUG-1996) to the EMBL/GenBank/DDBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=VN40;
 RX MEDLINE=95062197; PubMed=7972001;
 RA Tokita H., Okamoto H., Tsuda F., Song P., Nakata S., Chosa T.,
 RA Iizuka H., Mishiro S., Miyakawa Y., Mayumi M.,
 RA "Hepatitis C virus variants from Vietnam are classifiable into the
 RA seventh, eighth, and ninth major genetic groups.";
 RT Proc. Natl. Acad. Sci. U.S.A. 91:11022-11026(1994).
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA (BY SIMILARITY).
 EMBL; D88475; BAA13614.1; -.
 GO; GO:0016021; C:integral to membrane; IEA.
 DR GO; GO:0019028; C:viral capsid; IEA.
 DR GO; GO:0019031; C:viral envelope; IEA.
 DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR InterPro; IPR002522; HCV capsid.
 DR InterPro; IPR002521; HCV core.
 DR InterPro; IPR002519; HCV env.
 DR InterPro; IPR002531; HCV NS1.
 DR Pfam; PF01543; HCV_capsid; 1.
 DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV_env; 1.
 DR Pfam; PF01560; HCV_NS1; 1.
 DR Pfam; PF018602; HCV NS1; 1.
 KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
 KW Polypeptide; Transmembrane.
 FT NON_TER 414 414
 SQ SEQUENCE 414 AA; 44560 MW; 373D93269BEE3E2 CRC64;

Query Match 95.4%; Score 997; DB 12; Length 414;
 Best Local Similarity 95.3%; Pred. No. 9 7e-83;
 Matches 182; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Qy 1 MSLPKPQRKTRNTRPTDKFPGSGQIVGQVYLLPRGRGLGVATRKTSERSQPRG 60
 Db 1 MSLPKPQRKTRNTRPTDKFPGSGQIVGQVYLLPRGRGLGVATRKTSERSQPRG 60

Qy 61 RQQPIKARQPOGRHMAQPGVPMWPLYNGEGCWAGMLSPRSRPHGPNPDRRSNLG 120
 Db 61 RQQPIKARQPOGRHMAQPGVPMWPLYNGEGCWAGMLSPRSRPHGPNPDRRSNLG 120

Qy 121 KVIDLTGCGFADLMGYIPVVGAPLGGAVALAHLGVRAIEDGINYATGNLPGCSFSFLA 180
 Db 121 KVIDLTGCGFADLMGYIPVVGAPLGGAVALAHLGVRAIEDGINYATGNLPGCSFSFLA 180

Qy 181 LLSCLTPASA 191
 Db 181 LLSCLTPASA 191

RESULT 23

P89961 ID P89961 PRELIMINARY; PRT; 415 AA.
 AC B89961;
 DT 01-MAY-1997 (TREMBrel. 03, Created)
 DT 01-MAY-1997 (TREMBrel. 03, Last sequence update)
 DT 01-JUN-2003 (TREMBrel. 24, Last annotation update)
 DE Core, env and part of E2/NS1 (Genome polyprotein) (Fragment).
 DE Hepatitis C virus.
 OS Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=1103;
 RN SEQUENCE FROM N.A.
 [1]

RC STRAIN=VN540;
 RA Okamoto H.;
 RL Submitted (AUG-1996) to the EMBL/GenBank/DDBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=VN40;
 RX MEDLINE=95062197; PubMed=7972001;
 RA Tokita H., Okamoto H., Tsuda F., Song P., Nakata S., Chosa T.,
 RA Iizuka H., Mishiro S., Miyakawa Y., Mayumi M.,
 RA "Hepatitis C virus variants from Vietnam are classifiable into the
 RA seventh, eighth, and ninth major genetic groups.";
 RT Proc. Natl. Acad. Sci. U.S.A. 91:11022-11026(1994).
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 EMBL; D88474; BAA13623.1; -.
 GO; GO:0016021; C:integral to membrane; IEA.
 DR GO; GO:0019028; C:viral capsid; IEA.
 DR GO; GO:0019031; C:viral envelope; IEA.
 DR GO; GO:0003700; F:transcription factor activity; IEA.
 DR InterPro; IPR002522; HCV capsid.
 DR InterPro; IPR002521; HCV core.
 DR InterPro; IPR002519; HCV_env.
 DR InterPro; IPR002531; HTH_LysR.
 DR Pfam; PF01543; HCV_capsid; 1.
 DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV_env; 1.
 DR Pfam; PF01560; HCV_NS1; 1.
 DR Pfam; PS00044; HTH_LYSR_FAMILY; 1.
 KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
 KW Polypeptide; Transmembrane.
 FT NON_TER 415 415
 SQ SEQUENCE 415 AA; 44979 MW; B7C7222286C46A32 CRC64;

Query Match 95.4%; Score 997; DB 12; Length 415;
 Best Local Similarity 95.3%; Pred. No. 9 7e-83;
 Matches 182; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Qy 1 MSLPKPQRKTRNTRPTDKFPGSGQIVGQVYLLPRGRGLGVATRKTSERSQPRG 60
 Db 1 MSLPKPQRKTRNTRPTDKFPGSGQIVGQVYLLPRGRGLGVATRKTSERSQPRG 60

Qy 61 RQQPIKARQPOGRHMAQPGVPMWPLYNGEGCWAGMLSPRSRPHGPNPDRRSNLG 120
 Db 61 RQQPIKARQPOGRHMAQPGVPMWPLYNGEGCWAGMLSPRSRPHGPNPDRRSNLG 120

Qy 121 KVIDLTGCGFADLMGYIPVVGAPLGGAVALAHLGVRAIEDGINYATGNLPGCSFSFLA 180
 Db 121 KVIDLTGCGFADLMGYIPVVGAPLGGAVALAHLGVRAIEDGINYATGNLPGCSFSFLA 180

Qy 181 LLSCLTPASA 191
 Db 181 LLSCLTPASA 191

RESULT 24

P89964 ID P89964 PRELIMINARY; PRT; 415 AA.
 AC P89964;
 DT 01-MAY-1997 (TREMBrel. 03, Created)
 DT 01-MAY-1997 (TREMBrel. 03, Last sequence update)
 DT 01-JUN-2003 (TREMBrel. 24, Last annotation update)
 DE Core, env and part of E2/NS1 (Genome polyprotein) (Fragment).
 DE Hepatitis C virus.
 OS Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=1103;
 RN SEQUENCE FROM N.A.
 [1]

RC STRAIN=VN737;
 RC Okamoto H.;
 RA Submitted (AUG-1996) to the EMBL/GenBank/DDBJ databases.
 RIL [2]
 RN SEQUENCE FROM N.A.
 RP STRAIN=VN737;
 RX MEDLINE=95062197; PubMed=7972001;
 RA Tokita H., Okamoto H., Tsuda F., Song P., Nakata S., Chosa T.,
 RA Iizuka H., Mishiro S., Miyakawa Y., Mayumi M.;
 RT "hepatitis C virus variants from Vietnam are classifiable into the
 seventh, eighth, and ninth major genetic groups.";
 RT seven, eighth, and ninth major genetic groups.";
 RT proc. Natl. Acad. Sci. U.S.A. 91:11022-11026(1994).
 DR EMBL: D88477; BAM13626.1; -.
 GO: GO:001021; C:integral to membrane; IEA.
 GO: GO:001028; C:viral capsid; IEA.
 GO: GO:0019031; C:viral envelope; IEA.
 DR InterPro; IPR002522; HCV_capsid.
 DR InterPro; IPR002521; HCV_core.
 DR InterPro; IPR002519; HCV_envelope.
 DR InterPro; IPR002531; HCV_NS1.
 DR Pfam; PF0143; HCV_capsid; 1.
 DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV_envelope; 1.
 DR Pfam; PF01560; HCV_NS1; 1.
 DR Coat protein; EnvLop protein; Glycoprotein; Nonstructural protein;
 KW Polypeptide; Transmembrane.
 FT NON_TER 415 AA; 415 MW;
 SQ SEQUENCE 415 AA; 44771 MW;
 Query Match 95.4%; Score 997; DB 12; Length 415;
 Best Local Similarity 95.3%; Pred. No. 9_7e-83;
 Matches 182; Conservative 3; Mismatches 6; Indels 0; Gaps 0;
 Qy 1 MSLPKPORKTKRNTNRPTDKFPGQGQIVGGVYLPPRGRLGVRA
 Db 1 MSTLPKPORKTKRNTNRPTDKFPGQGQIVGGVYLPPRGRLGVRA
 Qy 61 RQPIPKARQPOGRHWAQPGVPMWPLQYNGEGCWAGMILSPRSRPHCPNDRRRSRLG 120
 Db 61 RQPIPKARQHQGRTWQPGVPMWPLQYNGEGCWAGMILSPRSRPHCPNDRRRSRLG 120
 Qy 121 KVDIITLGCFADLMGYIPVVGAPLGGVAALAHGVRAIEDGINYATGNLPGCSFSIFLLA 180
 Db 121 KVDIITLGCFADLMGYIPVVGAPLGGVAALAHGVRAIEDGINYATGNLPGCSFSIFLLA 180
 Qy 181 LISCLTPASA 191
 Db 181 LISCLTPASA 191

RESULT 25

P89965 PRELIMINARY; PRT; 415 AA.
 ID P89965
 AC P89965;
 DT 01-MAY-1997 (TREMBrel. 03, Last sequence update)
 DT 01-JUN-2003 (TREMBrel. 24, Last annotation update)
 DE Core, env and part of E2NS1 (Genome polyprotein) (Fragment).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Repliciviruses.
 OX NCBI_TaxID=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=VN843;
 RA Okamoto H.;
 RL Submitted (AUG-1996) to the EMBL/GenBank/DDBJ databases.
 RN [2]

RESULT 26

O92532 PRELIMINARY; PRT; 3015 AA.
 ID O92532
 AC O92532;
 DT 01-NOV-1998 (TREMBrel. 08, Last sequence update)
 DT 01-OCT-2003 (TREMBrel. 25, Last annotation update)
 DE Genome polyprotein.
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Repliciviruses.
 OX NCBI_TaxID=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=VN004;
 RA Okamoto H.;
 RL Submitted (MAR-1996) to the EMBL/GenBank/DDBJ databases.
 RN [2]

| | | | |
|----------|--|---|-----------|
| QY | 121 | KVDDDTGCGADLIMGVIPNVGAQPLGGVAAALAHGVRALEDGNYATGNLPGCSFSIFLA | 180 |
| Db | 121 | KVDDDTGCGADLIMGVIPNVGAQPLGGVAAALAHGVRALEDGNYATGNLPGCSFSIFLA | 180 |
| Qy | 181 | LISCLTPASA 191 | 180 |
| Db | 181 | LISCLTPASA 191 | 180 |
| | | | |
| RESULT | 27 | | |
| Q68153 | | | |
| ID | Q68153 | PRELIMINARY; | PRT: |
| AC | Q68153; | | 191 AA. |
| DT | 01-NOV-1996 | (TREMBLrel. 01, Created) | |
| DT | 01-NOV-1996 | (TREMBLrel. 01, Last sequence update) | |
| DT | 01-OCT-2003 | (TREMBLrel. 25, Last annotation update) | |
| DB | Genome polyprotein (Fragment). | | |
| GN | C. | | |
| OS | Hepatitis C virus. | | |
| OC | Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; | | |
| OC | Heptaviridae. | | |
| OX | NCBI_TAXID=11103; | | |
| [1] | | | |
| RP | SEQUENCE FROM N.A. | | |
| RC | STRAIN=Z4; | | |
| RX | MEDLINE=94336721; PubMed=8058787; | | |
| RC | STRAIN=Z4; | | |
| RX | MEDLINE=92279243; PubMed=1317578; | | |
| RA | Bukh J., Purcell R.H., Miller R.H.; | | |
| RA | Bukh J., Purcell R.H., Miller R.H.; | | |
| RA | "Sequence analysis of the 5' noncoding region of hepatitis C virus genotypes"; | | |
| RT | Proc. Natl. Acad. Sci. U.S.A. 91:82339-8243 (1994). | | |
| RN | [2] | | |
| RP | SEQUENCE FROM N.A. | | |
| RC | STRAIN=Z4; | | |
| RX | MEDLINE=93376778; PubMed=8396266; | | |
| RA | Bukh J., Purcell R.H., Miller R.H.; | | |
| RT | "At least 12 genotypes of hepatitis C virus predicted by sequence analysis of the putative 5' gene of isolates collected worldwide."; | | |
| RT | Proc. Natl. Acad. Sci. U.S.A. 90:8234-8238 (1993). | | |
| CC | - - SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA (BY SIMILARITY). | | |
| CC | PROTEIN C AND MRNA (BY SIMILARITY). | | |
| DR | EMBL; U10236; AAA1075.1; -. | | |
| DR | PIR; PQ0804; PQ0804. | | |
| DR | GO: 0019028; C:viral capsid; IEA. | | |
| GO | GO: 0005190; F:structural molecule activity; IEA. | | |
| InterPro | IPR002521; HCV core. | | |
| DR | InterPro; IPR002521; HCV core. | | |
| PFam | PF01543; HCV_capsid; 1. | | |
| PFam | PF01542; HCV_core; 1. | | |
| KW | Polyprotein. | | |
| FT | NON_TER | | |
| SEQUENCE | 191 | | |
| FT | SEQUENCE | 191 AA; | 20726 MW; |
| SQ | | 738DD912C66D8DC9 CRC64; | |

| | | | | |
|----|-------------|---|-----|--|
| Db | 61 | RQQPPIPKRQPGRSWMAQPGWPWLYGNEGCCWAGWLSPRSRPWGPNDPRRSNLG | 120 | RESULT 29 |
| Qy | 121 | KVIDLTLCGFAIDLGYIPVVGAPLGLSYAALAHGVRAIEDGINYATCNLPGCSFSIPLA | 180 | 081272 PRELIMINARY; PRT; 415 AA. |
| Db | 121 | KVIDLTLCGFAIDLGYIPVVGAPLGLSYAALAHGVRAIEDGINYATCNLPGCSFSIPLA | 180 | ID 081272 PRELIMINARY; PRT; 415 AA. |
| Qy | 181 | LISCLTPASA 191 | 180 | AC 081272 PRELIMINARY; PRT; 415 AA. |
| Db | 181 | LISCLTPASA 191 | 180 | DT 01-NOV-1995 (TREMBLrel. 01, Last sequence update) |
| Qy | 121 | KVIDLTLCGFAIDLGYIPVVGAPLGLSYAALAHGVRAIEDGINYATCNLPGCSFSIPLA | 180 | DT 01-NOV-1995 (TREMBLrel. 01, Last sequence update) |
| Db | 121 | KVIDLTLCGFAIDLGYIPVVGAPLGLSYAALAHGVRAIEDGINYATCNLPGCSFSIPLA | 180 | DT 01-NOV-1995 (TREMBLrel. 01, Last sequence update) |
| Qy | 181 | LISCLTPASA 191 | 180 | DT 01-NOV-1995 (TREMBLrel. 01, Last sequence update) |
| Db | 181 | LISCLTPASA 191 | 180 | DR Core, env and part of E2/NS1 (Genome polyprotein) (Fragment). |
| Qy | 121 | KVIDLTLCGFAIDLGYIPVVGAPLGLSYAALAHGVRAIEDGINYATCNLPGCSFSIPLA | 180 | OS Hepatitis C virus. |
| Db | 121 | KVIDLTLCGFAIDLGYIPVVGAPLGLSYAALAHGVRAIEDGINYATCNLPGCSFSIPLA | 180 | OC Hepacivirus. |
| Qy | 181 | LISCLTPASA 191 | 180 | OC Hepacivirus. |
| Db | 181 | LISCLTPASA 191 | 180 | OK NCBI_TaxId=1103; |
| Qy | 081315 | PRELIMINARY; PRT; 415 AA. | 180 | RN [1] |
| Db | 081315 | PRELIMINARY; PRT; 415 AA. | 180 | RP SEQUENCE FROM N.A. |
| Qy | 01-NOV-1995 | (TREMBLrel. 01, Last sequence update) | 180 | RC STRAIN=Th53; |
| Db | 01-NOV-1995 | (TREMBLrel. 01, Last sequence update) | 180 | DR MEDLINE=9505057; PubMed=7561773; |
| Qy | 01-JUN-2003 | (TREMBLrel. 24, Last annotation update) | 180 | RA Tokita H., Okamoto H., Luengrojanakul P., Vareesangthip K., Chainuvati T., Izuka H., Tsuda F., Miyakawa Y., Mayumi M.; |
| Db | 01-JUN-2003 | (TREMBLrel. 24, Last annotation update) | 180 | RA "Hepatitis C viruses variants from Vietnam are classifiable into the seventh, eighth, and ninth major genetic groups."; |
| Qy | 01-JUN-2003 | (TREMBLrel. 24, Last annotation update) | 180 | RA CC - SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA (BY SIMILARITY). |
| Db | 01-JUN-2003 | (TREMBLrel. 24, Last annotation update) | 180 | RA DR EMBL; D31971; BAA06739.1; -. |
| Qy | 01-JUN-2003 | (TREMBLrel. 24, Last annotation update) | 180 | RA DR GO; GO:0016021; C:integral to membrane; IEA. |
| Db | 01-JUN-2003 | (TREMBLrel. 24, Last annotation update) | 180 | RA DR GO; GO:0019028; C:viral capsid; IEA. |
| Qy | 01-JUN-2003 | (TREMBLrel. 24, Last annotation update) | 180 | RA DR GO; GO:001901; C:viral envelope; IEA. |
| Db | 01-JUN-2003 | (TREMBLrel. 24, Last annotation update) | 180 | RA DR GO; GO:000198; F:structural molecule activity; IEA. |
| Qy | 01-JUN-2003 | (TREMBLrel. 24, Last annotation update) | 180 | RA DR InterPro; IPR02522; HCV capsid. |
| Db | 01-JUN-2003 | (TREMBLrel. 24, Last annotation update) | 180 | RA DR InterPro; IPR02521; HCV core. |
| Qy | 01-JUN-2003 | (TREMBLrel. 24, Last annotation update) | 180 | RA DR InterPro; IPR02519; HCV env. |
| Db | 01-JUN-2003 | (TREMBLrel. 24, Last annotation update) | 180 | RA DR InterPro; IPR002531; HCV NS1. |
| Qy | 01-JUN-2003 | (TREMBLrel. 24, Last annotation update) | 180 | RA DR Pfam; PF01543; HCV capsid; 1. |
| Db | 01-JUN-2003 | (TREMBLrel. 24, Last annotation update) | 180 | RA DR Pfam; PF01542; HCV_core; 1. |
| Qy | 01-JUN-2003 | (TREMBLrel. 24, Last annotation update) | 180 | RA DR Pfam; PF01539; HCV_env; 1. |
| Db | 01-JUN-2003 | (TREMBLrel. 24, Last annotation update) | 180 | RA DR Pfam; PF01560; HCV_NS1; 1. |
| Qy | 01-JUN-2003 | (TREMBLrel. 24, Last annotation update) | 180 | RA KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein; Polyprotein; Transmembrane. |
| Db | 01-JUN-2003 | (TREMBLrel. 24, Last annotation update) | 180 | RA FT NON_TER 415 415 MW; 7P1B5354CD291728 CRC64; |
| Qy | 01-JUN-2003 | (TREMBLrel. 24, Last annotation update) | 180 | RA SQ SEQUENCE 415 AA; 44811 MW; |
| Db | 01-JUN-2003 | (TREMBLrel. 24, Last annotation update) | 180 | RA DR Query Match 95.2%; Score 995; DB 12; Length 415; Best Local Similarity 95.2%; Pred. No. 1.5e-82; Matches 181; Conservative 3; Mismatches 6; Indels 0; Gaps 0; |
| Qy | 1 | MSTLPKQKTKNTNRPTDKFPGGGQIVGSGVYLPRRGRLGVTRKTSERSOPRG | 60 | RA DR 1 MSTLPKQKTKNTNRPTDKFPGGGQIVGSGVYLPRRGRLGVTRKTSERSOPRG 60 |
| Db | 1 | MSTLPKQKTKNTNRPTDKFPGGGQIVGSGVYLPRRGRLGVTRKTSERSOPRG | 60 | RA DR 61 RQQPPIPKRQPGRSWMAQPGWPWLYGNEGCCWAGWLSPRSRPWGPNDPRRSNLG 60 |
| Qy | 61 | RQQPPIPKRQPGRSWMAQPGWPWLYGNEGCCWAGWLSPRSRPWGPNDPRRSNLG | 120 | RA DR 61 RQQPPIPKRQPGRSWMAQPGWPWLYGNEGCCWAGWLSPRSRPWGPNDPRRSNLG 60 |
| Db | 61 | RQQPPIPKRQPGRSWMAQPGWPWLYGNEGCCWAGWLSPRSRPWGPNDPRRSNLG | 120 | RA DR 121 KVIDLTLCGFAIDLGYIPVVGAPLGLSYAALAHGVRAIEDGINYATCNLPGCSFSIPLA 180 |
| Qy | 61 | RQQPPIPKRQPGRSWMAQPGWPWLYGNEGCCWAGWLSPRSRPWGPNDPRRSNLG | 120 | RA DR 121 KVIDLTLCGFAIDLGYIPVVGAPLGLSYAALAHGVRAIEDGINYATCNLPGCSFSIPLA 180 |
| Db | 61 | RQQPPIPKRQPGRSWMAQPGWPWLYGNEGCCWAGWLSPRSRPWGPNDPRRSNLG | 120 | RA DR 181 LISCLTPASA 191 |
| Qy | 121 | KVIDLTLCGFAIDLGYIPVVGAPLGLSYAALAHGVRAIEDGINYATCNLPGCSFSIPLA | 180 | RA DR 181 LISCLTPASA 191 |
| Db | 121 | KVIDLTLCGFAIDLGYIPVVGAPLGLSYAALAHGVRAIEDGINYATCNLPGCSFSIPLA | 180 | RA DR 181 LISCLTPASA 191 |
| Qy | 181 | LISCLTPASA 191 | 180 | RA DR RESULT 30 |
| Db | 181 | LISCLTPASA 191 | 180 | RA DR P87751 PRELIMINARY; PRT; 415 AA. |
| Qy | 181 | LISCLTPASA 191 | 180 | RA DR P87751 PRELIMINARY; PRT; 415 AA. |
| Db | 181 | LISCLTPASA 191 | 180 | RA DR DT 01-MAY-1997 (TREMBLrel. 03, Created) |

DT 01-MAY-1997 (TREMBLref: 03, Last sequence update)
 DT 01-JUN-2003 (TREMBLref: 24, Last annotation update)
 DE Core, env and part of E2/NS1 (Genome polyprotein) (Fragment).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=1103;
 RN [1]—
 RP SEQUENCE FROM N.A.
 RC STRAIN=VN235;
 RA Okamoto H.;
 RL Submitted [AUG-1996] to the EMBL/GenBank/DDBJ databases.
 RN [12]
 RP SEQUENCE FROM N.A.
 RC STRAIN=VN235;
 RX MEDLINE=95062197; PubMed=7972001;
 RA Tokita H., Okamoto H., Tsuda F., Song P., Nakata S., Chosa T.,
 RA Iizuka H., Mishiro S., Miyakawa Y., Mayumi M.;
 RT "Hepatitis C virus variants from Vietnam are classifiable into the
 seventh, eighth, and ninth major genetic groups.";
 Proc. Natl. Acad. Sci. U.S.A. 91:1022-1026(1994).
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA (BY SIMILARITY).
 DR EMBL; D88457; BA013616.1; -.
 DR GO; GO:0016021; C:integral to membrane; IEA.
 DR GO; GO:0019028; C:viral capsid; IEA.
 DR GO; GO:0019031; C:viral envelope; IEA.
 DR GO; GO:000198; F:structural molecule activity; IEA.
 DR Interpro; IPR002522; HCV_capsid.
 DR Interpro; IPR002521; HCV_core.
 DR Interpro; IPR002519; HCV_env.
 DR Interpro; IPR002531; HCV_NS1.
 DR Pfam; PF01543; HCV_capsid; 1.
 DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV_env; 1.
 DR Pfam; PF01560; HCV_NS1; 1.
 KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
 KW Polyprotein; Transmembrane.
 FT NON_TER 415 415 AA; 45301 MW;
 SQ SEQUENCE 415 AA; 39B2040A86517D1B CRC64;

Query Match 94.7%; Score 990; DB 12; Length 415;
 Best Local Similarity 94.2%; Ped. No. 4.2e-82;
 Matches 180; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

Matches 180; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY 1 MSTLPKQORKTKRNTNRNRPDKFPGGGQIVQSGVYLPRRGRRGRLGVATRKTSERSOPRG 60
 QY 61 RQQPQPKARRQRQTWQPGPMPLYNGECCWAGWLLSPRSRPRHGPNDRRSRNLG 120
 Db 61 RQQPQPKARRQRQTWQPGPMPLYNGECCWAGWLLSPRSRPRHGPNDRRSRNLG 120
 QY 121 KVIDLTGFGALMGMYIPVGAGPLGGVAALAHGVRAIEDGTNYATGNLPGCSFSIILLA 180
 Db 121 KVIDLTGFGADLGMGYIPVGAGPLGGVAALAHGVRAVEDGINYATGNLPGCSFSIILLA 180
 QY 181 LSCLTPASA 191
 Db 181 LSCLTPASA 191

Search completed: May 12, 2004, 09:15:59
 Job time : 43 secs

| OM protein - protein search, using SW model | | | | | | |
|---|---|---------------------------|--------------|----------------------------------|---------|----------------------|
| Run on: | May 12, 2004, 09:12:31 ; Search time 21 Seconds (without alignments) | | | 874.885 Million cell updates/sec | | |
| Title: | US-09-084-691B-206 | | | S40770 | | |
| Perfect score: | 1045 | | | 935 89.6 3033 1 GNWVQB | | |
| Sequence: | 1 MSLUPKPQRKVRNTRRPT.....CSFSIPLLLALISCLTPASA 191 | | | 919 87.9 178 2 PS0388 | | |
| Scoring table: | BL0SUM62 | | | 918 87.9 322 2 JN2065 | | |
| Scoring table: | Gapext 10.0 , Gapext 0.5 | | | 893 85.5 3014 1 JC5620 | | |
| Searched: | 283366 seqs, 96191526 residues | | | 841.9 124 2 S41360 | | |
| Post-processing: | Minimum Match 0% | | | 631.5 60.8 123 2 S41361 | | |
| Database : | Maximum Match 100% Listing first 90 summaries | | | 631.5 60.4 256 2 PQ0393 | | |
| Result No. | Score | Query Match Length | DB ID | 631.5 60.4 256 2 PQ0393 | | |
| SUMMARIES | | | | | | |
| 1 | 977 | 93.5 | 3011 | 1 | S40770 | genome polyprotein |
| 2 | 976 | 93.4 | 3011 | 1 | GNWVQB | genome polyprotein |
| 3 | 971 | 92.9 | 441 | 2 | S12707 | genome polyprotein |
| 4 | 969 | 92.7 | 782 | 2 | S18031 | genome polyprotein |
| 5 | 969 | 92.7 | 787 | 2 | PN0677 | hypothetical protein |
| 6 | 968 | 92.6 | 3011 | 1 | GNWVQC3 | genome polyprotein |
| 7 | 965 | 92.3 | 3010 | 1 | GNWVTC | genome polyprotein |
| 8 | 965 | 92.3 | 3010 | 1 | GNWVJC | genome polyprotein |
| 9 | 962 | 92.1 | 550 | 2 | JH0711 | genome polyprotein |
| 10 | 961 | 92.0 | 2 | JQ1584 | 640 | genome polyprotein |
| 11 | 959 | 91.8 | 189 | 2 | S32740 | polyprotein - hepa |
| 12 | 958 | 91.7 | 782 | 2 | S19876 | genome polyprotein |
| 13 | 957 | 91.6 | 513 | 2 | PC1284 | genome polyprotein |
| 14 | 957 | 91.6 | 782 | 2 | S19875 | genome polyprotein |
| 15 | 957 | 91.6 | 876 | 2 | PC2219 | polypeptide - hepa |
| 16 | 956 | 91.5 | 3010 | 1 | S18030 | genome polyprotein |
| 17 | 956 | 91.5 | 3010 | 1 | GNWVW | genome polyprotein |
| 18 | 952 | 91.1 | 520 | 2 | JQ1225 | polyprotein - hepa |
| 19 | 951 | 91.0 | 411 | 2 | PC2060 | genome polyprotein |
| 20 | 948 | 90.7 | 3010 | 1 | A45573 | genome polyprotein |
| 21 | 947 | 90.6 | 369 | 2 | S21471 | genome polyprotein |
| 22 | 947 | 90.6 | 523 | 2 | JQ1926 | polyprotein - hepa |
| 23 | 947 | 90.6 | 782 | 2 | S18032 | genome polyprotein |
| 24 | 946 | 90.5 | 492 | 2 | S41288 | genome polyprotein |
| 25 | 944 | 90.3 | 411 | 2 | PC2061 | genome polyprotein |
| 26 | 941 | 90.0 | 874 | 2 | JQ0883 | genome polyprotein |
| 27 | 940 | 90.0 | 513 | 2 | A44150 | structural protein |
| 28 | 939 | 89.9 | 874 | 2 | JQ0881 | genome polyprotein |
| 29 | 939 | 89.9 | 3033 | 1 | JQ1303 | genome polyprotein |
| RESULT 1 | | | | | | |
| ALIGNMENTS | | | | | | |
| N: Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructural protein NS4a); nonstructural protein NS4b; nonstructural protein NS5 | | | | | | |
| C: Species: hepatitis C virus | | | | | | |
| C: Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001 | | | | | | |

C;Accession: S40770; PID:PC1285
A;Residues: 1-311 <OKA>
A;Cross-references: EMBL:D10749; NID:9221586; PIDN:BAA01582.1; PID:9221587
R;Okamoto, H.; Okada, S.; Sugiyama, Y.; Yotsuji, S.; Tanaka, T.; Yoshizawa, H.; Tsuda, Jpn. J. Exp. Med. 60, 167-177, 1990
A;Title: The 5'-terminal sequence of the hepatitis C virus genome.
A;Reference number: PC1284; MUID:91013116; PMID:2170712
A;Accession: PC1285
A;Molecule type: genomic RNA
A;Residues: 1-513 <OKZ>
A;Cross-references: GB:D00831; NID:9221511; PIDN:BAA00705.1; PID:9221512
A;Experimental source: isolate HC-JI
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural
C;1-115/Product: capsid protein C #status predicted <CP>
F;1-15/Product: capsid protein C #status predicted <CP>
F;1-191/Product: envelope protein M #status predicted <EPM>
F;192-389/Product: major envelope protein E #status predicted <MEB>
F;390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F;1007-1615/Product: hepacivirin #status predicted <NS3>
F;1230-1237/Region: nucleotide-binding motif A (P-loop)
F;1230-1237/Region: nucleotide-binding motif A (P-loop)
F;1312-1317/Region: DEXH motif
F;1316-1319/Region: DEXH motif
F;1862-Product: nonstructural protein NS4b #status predicted <NS4>
F;2014-3011/Region: nonstructural protein NS5 #status predicted <NS5>
F;2016-2017/Region: nonstructural protein NS5 #status predicted <NS5>
F;2014-3011/Region: nonstructural protein NS5 #status predicted <NS5>
F;2016-2017/Region: nonstructural protein NS5 #status predicted <NS5>
F;1310-1319/Region: nucleotide-binding motif A (P-loop)
F;1616-1862/Product: nonstructural protein NS4a #status predicted <NS4a>
F;1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>
F;1312-1317/Region: nucleotide-binding motif A (P-loop)
F;1312-1317/Region: nucleotide-binding motif A (P-loop)
Query Match 93.5%; Score 977; DB 1; Length 3011;
Best Local Similarity 92.1%; Pred. No. 1.1e-74;
Matches 176; Conservative 6; Mismatches 9; Indels 0; Gaps 0;
Db 1 MSTLPKPKRKTKRNTNRPRQDKVFPGGQIVGVYVLLPRRGPRIGVATRKTSERSQPG 60
1 MSTLPKPKRKTKRNTNRPRQDKVFPGGQIVGVYVLLPRRGPRIGVATRKTSERSQPG 60
Qy 61 RQPIPKARQPOGRHWAQPGVPLGEGCGWAGWILSPRSSRPHQPNPDRRSNLG 120
Db 61 RQPIPKARQPOGRHWAQPGVPLGEGCGWAGWILSPRSSRPHQPNPDRRSNLG 120
Qy 121 KVIDTLCGFADLMGYIPVVGAPLGVAALAHGVRAEDGINYATGNLPGCSFSIFLA 180
Db 121 KVIDTLCGFADLMGYIPVVGAPLGVAALAHGVRAEDGINYATGNLPGCSFSIFLA 180
Qy 181 LLSCLTTPASA 191
Db 181 LLSCLTTPASA 191

RESULT 2

GNWUCH
genome polyprotein - hepatitis C virus (strain H)
N;Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructural
protein NS4a); nonstructural protein NS4b; nonstructural protein NS5
C;Species: hepatitis C virus
A;Note: host Homo sapiens (man)
C;Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001
C;Accession: A36814; A41545
R;Inchausepe, G.; Zebadee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.
submitted to Genbank, July 1992
A;Description: Genomic structure of the human prototype strain H of hepatitis C virus: c
A;Reference number: A36814
A;Accession: A36814
A;Molecule type: genomic RNA
A;Residues: 1-3011 <INC>
A;Cross-references: GB:M67463; NID:9329737; PIDN:AAA45534.1; PID:9329738
R;Inchausepe, G.; Zebadee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.
Proc. Natl. Acad. Sci. U.S.A. 88, 10282-10296, 1991
A;Title: Genomic structure of the human prototype strain H of hepatitis C virus: compar

A;Reference number: A41546; MUID:92052256; PMID:1658800
A;Contents: annotation
A;Note: neither amino acid nor nucleotide sequence is given
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural
C;1-115/Product: capsid protein C #status predicted <CP>
F;116-191/Product: envelope protein M #status predicted <EPM>
F;192-389/Product: major envelope protein E #status predicted <MEB>
F;390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F;1007-1615/Product: hepacivirin #status predicted <NS3>
F;1230-1237/Region: nucleotide-binding motif A (P-loop)
F;1312-1317/Region: nucleotide-binding motif B
F;1316-1319/Region: DEXH motif
F;1616-1862/Product: nonstructural protein NS4b #status predicted <NS4>
F;1863-2013/Product: nonstructural protein NS5 #status predicted <NS5>
F;2014-3011/Region: nonstructural protein NS5 #status predicted <NS5>
F;2016-2017/Region: nonstructural protein NS5 #status predicted <NS5>
F;2014-3011/Region: nonstructural protein NS5 #status predicted <NS5>
F;2016-2017/Region: nonstructural protein NS5 #status predicted <NS5>
Query Match 93.4%; Score 976; DB 1; Length 3011;
Best Local Similarity 92.7%; Pred. No. 2.1e-75;
Matches 177; Conservative 5; Mismatches 9; Indels 0; Gaps 0;
Db 1 MSTLPKPKRKTKRNTNRPRQDKVFPGGQIVGVYVLLPRRGPRIGVATRKTSERSQPG 60
1 MSTLPKPKRKTKRNTNRPRQDKVFPGGQIVGVYVLLPRRGPRIGVATRKTSERSQPG 60
Qy 61 RQPIPKARQPOGRHWAQPGVPLGEGCGWAGWILSPRSSRPHQPNPDRRSNLG 120
Db 61 RQPIPKARQPOGRHWAQPGVPLGEGCGWAGWILSPRSSRPHQPNPDRRSNLG 120
Qy 121 KVIDTLCGFADLMGYIPVVGAPLGVAALAHGVRAEDGINYATGNLPGCSFSIFLA 180
Db 121 KVIDTLCGFADLMGYIPVVGAPLGVAALAHGVRAEDGINYATGNLPGCSFSIFLA 180
Qy 181 LLSCLTTPASA 191
Db 181 LLSCLTTPASA 191

RESULT 3

S2707
genome polyprotein - hepatitis C virus (fragment)
N;Contains: core protein; envelope protein
C;Species: hepatitis C virus
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 17-Nov-2000
C;Accession: S1207
R;Takeuchi, K.; Kubo, Y.; Boomar, S.; Watanabe, Y.; Katayama, T.; Choo, Q.L.; Kuo, G.; I
Nucleic Acid Res. 18, 4626, 1990
A;Title: Nucleotide sequence of core and envelope genes of the hepatitis C virus genome c
A;Reference number: S1207; MUID:90356432; PMID:211749
A;Accession: S1207
A;Molecule type: genomic RNA
A;Residues: 1-441 <TAK>
A;Cross-references: EMBL:D00574; NID:9221656; PIDN:BAA00452.1; PID:9221657
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: polyprotein
Query Match 92.9%; Score 971; DB 2; Length 441;
Best Local Similarity 92.7%; Pred. No. 9.1e-75;
Matches 177; Conservative 5; Mismatches 9; Indels 0; Gaps 0;
Db 1 MSTLPKPKRKTKRNTNRPRQDKVFPGGQIVGVYVLLPRRGPRIGVATRKTSERSQPG 60
1 MSTLPKPKRKTKRNTNRPRQDKVFPGGQIVGVYVLLPRRGPRIGVATRKTSERSQPG 60
Qy 61 RQPIPKARQPOGRHWAQPGVPLGEGCGWAGWILSPRSSRPHQPNPDRRSNLG 120
Db 61 RQPIPKARQPOGRHWAQPGVPLGEGCGWAGWILSPRSSRPHQPNPDRRSNLG 120
Qy 121 KVIDTLCGFADLMGYIPVVGAPLGVAALAHGVRAEDGINYATGNLPGCSFSIFLA 180
Db 121 KVIDTLCGFADLMGYIPVVGAPLGVAALAHGVRAEDGINYATGNLPGCSFSIFLA 180

| | | | |
|--|-----|---|-----|
| Db | 181 | LISCLTPASA | 191 |
| Qy | 121 | KVDTLTCGFADIMGYIPVGAPGGVAALAHGSRATIEDGINTATGNLPGCSFSIFLLA | 180 |
| Db | 121 | KVDTLTCGFADIMGYIPVGAPGGVAALAHGSRATIEDGINTATGNLPGCSFSIFLLA | 180 |
| Qy | 181 | LISCLTPASA | 191 |
| RESULT | 7 | | |
| GNWTC | | | |
| N;Contains: capsid protein C; envelope protein M; major envelope protein E; nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5 | | | |
| C;Species: hepatitis C virus | | | |
| C;Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 19-Jan-2001 | | | |
| C;Accession: A38465 | | | |
| A;Title: Structure and organization of the hepatitis C virus genome isolated from human amizawa, A.; Mori, C.; Fuke, I.; Manabe, S.; Murakami, S.; Fujita, J.; Onishi, E.; | | | |
| A;Reference: A38465 | | | |
| A;Accession: A38465 | | | |
| A;Molecule type: genomic RNA | | | |
| A;Residues: 1-3010 <RAX> | | | |
| A;Cross-references: EMBL:MS8335; NID:9329770; PID:AAA72945.1; PID:9329771 | | | |
| C;Superfamily: hepatitis C virus genome polyprotein | | | |
| C;Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural | | | |
| F;116-191/Product: envelope protein M #status predicted <PEM> | | | |
| F;191-389/Product: major envelope protein E #status predicted <MEB> | | | |
| F;390-729/Product: nonstructural protein NS1 #status predicted <NS1> | | | |
| F;730-1065/Product: nonstructural protein NS2 #status predicted <NS2> | | | |
| F;1007-1615/Product: hepacivirin #status predicted <NS3> | | | |
| F;1230-1237/Region: nucleotide-binding motif A (P-loop) | | | |
| F;1312-1317/Region: nucleotide-binding motif B | | | |
| F;1316-1319/Region: DEXH motif | | | |
| F;1616-1862/Product: nonstructural protein NS4a #status predicted <NA> | | | |
| F;1863-2033/Product: nonstructural protein NS4b #status predicted <NKB> | | | |
| F;2014-3010/Product: nonstructural protein NS5 #status predicted <NS5> | | | |
| F;195, 209, 234, 250, 305, 325, 417, 423, 430, 448, 532, 540, 556, 576, 623, 645, 1213, 1255, 2041, 2077, 22 | | | |
| Query Match 92.3% Score 965: DB 1; Length 3010; | | | |
| Best Local Similarity 92.1%; Pred. No. 1.8e-73; 10; Indels 0; Gaps 0; | | | |
| Matches 176; Conservative 5; Mismatches 0; | | | |
| Qy | 1 | MSTLPKQRTKENTNRPTDVKPGQQIVGGYILPRRGPRGRATRKTSERSQPRG | 60 |
| Db | 1 | MSTNPKFQRTKENTNRPTDVKPGQQIVGGYILPRRGPRGRATRKTSERSQPRG | 60 |
| Qy | 61 | RRQPPIPKARQPOGHWAQOPGYPWMLYGNQCGWAGWLSPRSRSHWGPMDPRRSRLG | 120 |
| Db | 61 | RRQPPIPKARQPOGHWAQOPGYPWMLYGNQCGWAGWLSPRSRSHWGPMDPRRSRLG | 120 |
| Qy | 121 | KVDTLTCGFADIMGYIPVGAPGGVAALAHGSRATIEDGINTATGNLPGCSFSIFLLA | 180 |
| Db | 121 | KVDTLTCGFADIMGYIPVGAPGGVAALAHGSRATIEDGINTATGNLPGCSFSIFLLA | 180 |
| Qy | 181 | LISCLTPASA | 191 |
| Db | 181 | LISCLTPASA | 191 |

proc. Natl. Acad. Sci. U.S.A. 87, 9524-9528, 1990
 Article: Molecular cloning of the human hepatitis C virus genome from Japanese Patients v
 A;Reference number: A39253; MUID:91088550; PMID:2175903
 A;Accession: AJ9253
 A;Molecule type: genomic RNA
 A;Cross-references: GB:D00208; NID:9221610; PID:BAA1433.1; PID:9221611
 RIKATO, N.; OHKoshi, S.; SHIMODORO, K.
 Proc. Jpn. Acad. 65B, 219-223, 1989
 A;Title: Japanese isolates of the non-A, non-B hepatitis viral genome show sequence vari
 A;Reference number: PS0085
 A;Accession: PS0086
 A;Molecule type: genomic RNA
 A;Residues: 2650-2707 <RAS>
 A;Experimental source: Japanese isolate
 C;Comment: The cleavage sites of this polyprotein have not been determined.
 C;Superfamily: hepatitis C virus genome polyprotein
 C;Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; polyprotein; serine
 F;2-15/ product: capsid protein C #status predicted <CPC>
 F;116-191/ product: envelope protein M #status predicted <EPM>
 F;192-389/ product: major envelope protein E #status predicted <MBB>
 F;397-725/ product: nonstructural protein NS1 #status predicted <NS1>
 F;730-1006/ product: nonstructural protein NS2 #status predicted <NS2>
 F;1007-1615/ product: hepacivirin #status predicted <NS3>
 F;1230-1237/Region: nucleotide-binding motif A (P-loop)
 F;1311-1311/Region: DEXH motif
 F;1616-1862/ Product: nonstructural protein NS4a #status predicted <NS4A>
 F;1863-2013/ Product: nonstructural protein NS4b #status predicted <NS4B>
 F;1961-209,234,250,305,325,417-423,430,448,532,556,576,623,655,1213,1255,2041,2077,2240,2:
 Query Match 92.3%; Score 965; DB 1; Length 3010;
 Best Local Similarity 92.1%; Pred. No. 1. 8e-73; 5; Mismatches 10; Indels 0; Gaps 0;
 Matches 176; Conservative 5;
 Qy 1 MSTLPKQORKIGNTNRRPTDVKPFCGCGTQVGVVILPRRGRLGRATRKTSERQPRG 60
 Db 1 M\$TLPKQORKIGNTNRRPTDVKPFCGCGTQVGVVILPRRGRLGRATRKTSERQPRG 60
 Qy 61 RRPQPKPQRQGPGRHNAQPGTPWPLKGNEGCCWAGHMLSPRSRSPRHGPNPDRRRSNLQ 120
 Db 61 RRPQPKPQRGPGRHNAQPGTPWPLKGNEGCCWAGHMLSPRSRSPRHGPNPDRRRSNLQ 120
 Qy 121 KVDTITCGFAIDLGMVYIPWVGAPLGCVAAALAHGVRAEDGVNATGTLPGCSFSFLA 180
 Db 121 KVDTITCGFAIDLGMVYIPWVGAPLGCVAAALAHGVRAEDGVNATGTLPGCSFSFLA 180
 Qy 181 LLSCLTTPASA 191
 Db 181 LLSCLTTPASA 191
 RESULT 9
 JR0711 genome polyprotein - hepatitis C virus (strain PR1) (fragments)
 N;Contains: envelop protein E1; envelope protein E2; nonstructural protein NS1; nonstruc
 C;Species: hepatitis C virus
 C;Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 17-Nov-2000
 C;Accession: JR0711
 R,Liu, K.; Hu, Z.; Li, H.; Prince, A.M.; Inchauspe, G.
 Gene 114, 245-250, 1992
 A;Title: Genomic typing of hepatitis C viruses present in China.
 A;Reference number: JR0711; MUID:92290283; PMID:1318245
 A;Accession: JR0711
 A;Molecule type: genomic RNA
 A;Residues: 1-550 <LIU>
 A;Cross-references: GB:MT4888; GB:MT4889
 A;Note: the nucleotide sequence is not complete
 A;Note: translation of the nucleotide sequence is not complete
 C;Superfamily: hepatitis C virus genome polyprotein
 C;Keywords: envelope protein; glycoprotein; nonstructural protein; nucleocapsid; polyprot
 F;1-190/ Product: nucleocapsid protein C #status predicted <CPC>

F;191-380/Product: envelope protein E1 #status predicted <EP1>
 F;381-514/Product: envelope protein E2 or nonstructural protein NS1 (fragment) #status p
 F;515-550/Product: nonstructural protein NS5 (fragment) #status predicted <NS5>
 F;196,233,250,305,416,422,429,447/Binding site: carbohydrate (Asn) (covalent) #status pr

Query Match
 Best Local Similarity 92.1%; Score 962; DB 2; Length 550;
 Matches 175; Conservative 6; Mismatches 10; Indels 0; Gaps 0;

QY 1 MSTLPKQORKTIGENTNRPTDKFPGGGQIVQGVVYLPRRGRLGVRAKTKRSERSOPRG 60

Db 1 MSTNPKPQRTKNTNRPTDKFPGGGQIVQGVVYLPRRGRLGVRAKTKRSERSOPRG 60

QY 61 RROPIPKARQPOCRHWAQPGYPMWPLYNGEGCWAGWILSPRSRSPHGPNDRRSRNLG 120

Db 61 RROPIPKARRPEGRTWAAQPGYPMWPLYNGEGCWAGWILSPRSRSPHGPNDRRSRNLG 120

QY 121 KVIDTLCGFAIDLGYIPLVVGAPLGVAALAHGVRAIEDGINYATGNLPGCSFSFLIA 180

Db 121 KVIDTLCGFAIDLGYIPLVVGAPLGVAALAHGVRAIEDGINYATGNLPGCSFSFLIA 180

QY 181 LSCLTTPASA 191

Db 181 LSCLTTPASA 191

RESULT 10

JQ1584 Genome polyprotein - hepatitis C virus (strain U.K.) (fragment)

C;Contains: core protein C; envelope protein E1; envelope protein

C;Keywords: hepatitis C virus

C;Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 17-Nov-2000

C;Accession: JQ1584

R;Kumar, U.; Cheng, D.; Thomas, H.; Monjardino, J.

J; Gen. Virol. 73, 1521-1525, 1992

A;Title: Cloning and sequencing of the structural region and expression of putative core

A;Reference number: JQ1584; MUID:92300349; PMID:1318944

A;Accession: JQ1584

A;Molecule type: genomic RNA

A;Residues: 1-640 <RDN>

A;Cross-references: GB:X84079; NID:9643119; PIDN:CAA58888.1; PID:9643120

C;Superfamily: hepatitis C virus genome polyprotein

C;Keywords: core protein; envelope protein; glycoprotein; nonstructural protein; polypro

F;1-191/Product: core protein C #status predicted <CPC>

F;192-389/Product: envelope protein E1 #status predicted <ERI>

F;390-640/Product: envelope protein E2 and nonstructural protein NS1 #status predicted <

F;196,209,234,305,417,430,448-476,540,556,576-623/Binding site: carbohydrate (Asn) (cova

F;374-782/Product: nonstructural protein 2 (fragment) #status predicted <MAT4>

F;734-782/Product: nonstructural protein 2 (fragment) #status predicted <MAT4>

RESULT 12

S19876 Genome polyprotein - hepatitis C virus (isolate JK5) (fragment)

N;Contains: core protein; envelope protein 1; nonstructural protein 2; NS1/E2 protein

C;Species: hepatitis C virus

A;Variety: isolate JK5

C;Keywords: hepatitis C virus

C;Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 17-Nov-2000

C;Accession: S19876

R;Honda, M.; Kaneko, S.; Masashi, U.; Kobayashi, K.; Murakami, S.

Submitted to the EMBL Data Library, September 1991

A;Description: Sequence analysis of putative structural regions of Hepatitis C Virus iso

A;Reference number: S18029

A;Accession: S19876

A;Molecule type: genomic RNA

A;Residue: 1-782 <RDN>

A;Cross-references: EMBL:X61595; NID:959486; PIDN:CAA43792.1; PID:959487

A;Experimental source: isolate JK5

C;Superfamily: hepatitis C virus genome polyprotein

C;Keywords: capsid protein; core protein; envelope protein; glycoprotein; nonstructural

F;1-191/Product: core protein #status predicted <MAT1>

F;192-389/Product: envelope protein 1 #status predicted <MAT1>

F;384-733/Product: NS1/E2 protein #status predicted <MAT3>

F;734-782/Product: nonstructural protein 2 (fragment) #status predicted <MAT4>

Query Match 91.7%; Score 958; DB 2; Length 782;
 Best Local Similarity 91.6%; Pred. No. 2e-73; Matches 175; Conservative 4; Mismatches 12; Indels 0; Gaps 0;

QY 1 MSTLPKQORKTIGENTNRPTDKFPGGGQIVQGVVYLPRRGRLGVRAKTKRSERSOPRG 60

Db 1 MSTNPKPQRTKNTNRPTDKFPGGGQIVQGVVYLPRRGRLGVRAKTKRSERSOPRG 60

QY 61 RROPIPKARQPOCRHWAQPGYPMWPLYNGEGCWAGWILSPRSRSPHGPNDRRSRNLG 120

Db 61 RROPIPKARRPEGRTWAAQPGYPMWPLYNGEGCWAGWILSPRSRSPHGPNDRRSRNLG 120

QY 121 KVIDTLCGFAIDLGYIPLVVGAPLGVAALAHGVRAIEDGINYATGNLPGCSFSFLIA 180

Db 121 KVIDTLCGFAIDLGYIPLVVGAPLGVAALAHGVRAIEDGINYATGNLPGCSFSFLIA 180

QY 181 LSCLTTPASA 191

Db 181 LSCLTTPASA 191

RESULT 11

S32740 Polyprotein - hepatitis C virus (isolate Russian) (fragment)

N;Contains: capsid protein C; envelope protein M

C;Species: hepatitis C virus
 C;Name: isolate JK1
 C;Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 23-Mar-2001
 C;Accession: S18030; S33570; R48332; S88029
 R;Honda, M.; Kaneko, S.; Ueda, M.; Kobayashi, K.; Murakami, S.
 submitted to the EMBL Data Library, September 1991
 A;Description: A whole genome of hepatitis C virus cDNA was isolated from a single patient.
 A;Reference number: S18028
 A;Accession: S18030
 A;Molecule type: genomic RNA
 A;Residues: 1-3010 <HON>
 A;Cross-references: EMBL:X61596; NID:959478; PIDN:CAA43793.1; PID:959479
 A;Experimental source: isolate JK1 from an individual
 R;Honda, M.; Kaneko, S.; Ueda, M.; Kobayashi, K.; Murakami, S.
 Arch. Virol. 128, 163-169, 1993
 A;Title: Sequence analysis of putative structural regions of hepatitis C virus isolated
 A;Reference number: A48332; PMID:93119270; PMID:8380322
 A;Accession: S33570
 A;Molecule type: genomic RNA
 A;Length: 1-547, 'T', 549-621, 'V', 623-624, 'S', 626-652, 'D', 655-761, 'T', 763-782 <HOW>
 A;Note: this sequence is inconsistent with the nucleotide translation
 A;Note: the authors translated the codon AGG for residue 43 as Pro, TGG for residue 320 as Trp, and TTC for residue 771 as Ser
 A;Note: sequence extracted from NCBI backbone (NBRIN:121747, NCBIPI:121748)
 C;Keywords: hepatitis C virus genome; polyprotein
 C;Keywords: AMP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin
 C;Keywords: 2'-15'/Product: capsid protein C #status predicted <CP>
 F;116-111/Product: envelope protein M #status predicted <EPM>
 F;192-389/Product: major envelope protein E #status predicted <MEB>
 F;390-729/Product: nonstructural protein NS1 #status predicted <NS1>
 F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
 F;1007-1615/Product: hepacivirin #status predicted <NS3>
 F;1230-1237/Region: nucleotide-binding motif A (P-loop)
 F;1312-1317/Region: nucleotide-binding motif B
 F;1316-1319/Region: DEXH motif
 F;1616-1862/Product: nonstructural protein NS4a #status predicted <NS4a>
 F;1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>
 F;1910-2010/Product: nonstructural protein NS5 #status predicted <NS5>
 F;196,209,234,250,305,417,423,448,532,540,556,576,623,645/Binding Site: carbohydrate (As
 F;196,209,234,250,305,417,423,448,532,540,556,576,623,645/Binding Site: carbohydrate (As
 Query Match 91 %; Score 956; DB 1; Length 3010;
 Best Local Similarity 91.6%; Pred. No. 1e-72; Matches 174; Conservative 4; Mismatches 12; Indels 0; Gaps 0;
 QY 1 MSTLDPKPKQRTKRNTRRRPTDVKEPGGQIVGGVILPRGRPLGRATRKTSERSQPRG 60
 D^b 1 MSTNPKPKQRTKRNTRRRPTDVKEPGGQIVGGVILPRGRPLGRATRKTSERSQPRG 60
 Q 61 RRQPQPKARQPOGRWAQPGYFWPLWGNESCGWAGMLSLSPRSRPRWGPNDPRRSNLG 120
 D^b 61 RRQPQPKARQPOGRWAQPGYFWPLWGNESCGWAGMLSLSPRSRPRWGPNDPRRSNLG 120
 QY 121 KVIDTUTCGFADLMSYTPVVGAPLGGAALAHGVALEDGINYATGNLPGCSFSFLA 180
 Db 121 KVIDTUTCGFADLMSYTPVVGAPLGGAALAHGVALEDGINYATGNLPGCSFSFLA 180
 QY 181 LLSCLTTPAS 190
 Db 181 LLSCLTTPVS 190
 181 LLSCLTTPVS 190

RESULT 17

genome polyprotein - hepatitis C virus (strain Taiwan)
N;Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructural protein NS4b); nonstructural protein NS5
C;Species: hepatitis C virus
A;Note: host: Homo sapiens (man)
C;Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001
C;Accession: A40244
R;Chen, P.J.; Lin, M.H.; Tai, K.F.; Liu, P.C.; Lin, C.J.; Chen, D.S.
Virology 188: 102-113, 1992

Matche

F;#residues: 419; #glycoprotein; #nucleotide; nucleotide binding; p-loop; polyproline; Berlin
 F;2-115/Product: capsid protein C #status predicted <CPC>
 F;116-191/Product: envelope protein M #status predicted <BN>
 F;192-389/Product: major envelope protein E #status predicted <MEE>
 F;390-729/Product: nonstructural protein NS1 #status predicted <NS1>
 F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
 F;107-1615/Product: hepacitinrin #status predicted <NS3>
 F;1230-1237/Region: nucleotide-binding motif A (P-loop)
 F;1312-1317/Region: nucleotide-binding motif B
 F;1316-1319/Region: DEXH motif
 F;161-186/Region: nonstructural protein NS4a #status predicted <NSA>
 F;1863-2013/Product: nonstructural protein NS4b #status predicted <NSB>
 F;2014-3010/Product: nonstructural protein NS5 #status Predicted <NS5>
 F;196,209,234,250,305,417,423,448,532,540,556,576,623,645/Binding site: carbohydrate (As
 Query Match 91.5%; Score 956; DB 1; Length 3010;
 Best Local Similarity 91.6%; Pred. No. 1e-72; 12; Indels 0; Gaps 0;
 Matches 174; Conservative 4; Mismatches

RESULT 18

Magazine

卷之三

F-1230-12

1923-385

C:\Keywords

A;Residue

A;FILE:
A:REFER

RESULT 19

Db 61 RQPIPKARRPEGRAGAWAQPGYWPMLYNEGGWAGWNLSPRSRSPWGPDTRRSRNLG 120

Qy 121 KVIDTUTCGFADLGMYIPWVGAPLGGAVALAHLGVRAIEDGINYATGNLPGCSFSFLA 180

Db 121 KVIDTUTCGFADLGMYIPWVGAPLGGAVALAHLGVRAIEDGINYATGNLPGCSFSFLA 180

Qy 181 LNSCLTPASA 191

Db 181 LNSCLTPASA 191

RESULT 20

Db 61 RQPIPKARRPEGRAGAWAQPGYWPMLYNEGGWAGWNLSPRSRSPWGPDTRRSRNLG 120

Qy 116-191/Prd: capsid protein C #status predicted <CPC>

C;Species: hepatitis C virus

C;Date: 14-Jul-1994 #sequence_revision 14-Jul-1994 #text_change 17-Nov-2000

C;Accession: PC2060

F;J.S.; Virtsitski, L.; Tong, S.P.; Trepo, C.

F;em. Biophys. Res. Commun. 199, 1474-1481, 1994

A;Title: Identification of the third major genotype of hepatitis C virus in France.

A;Reference number: PC2060; MUID:94197744; PMID:8147893

A;Accession: PC2060

A;Molecule type: mRNA

A;Residues: 1-411 <LIJ>

C;Superfamily: hepatitis C virus genome polyprotein

C;Keywords: ATP; capsid protein; envelope protein; glycoprotein; nonstructural protein; F;192-383/Product: envelope protein B1 #status predicted <SPS> F;384-411/Product: nonstructural protein B2/NS1 #status predicted <NPE> F;196, 209, 234, 305, 325/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match Best Local Similarity 91.0%; Score 951; DB 2; Length 411; Matches 172; Conservative 9; Mismatches 10; Indels 0; Gaps 0;

Qy 1 MSTLPKPKRKTKRNTNRPTDKFPGGQIVGVYLPRRGRLGRVTRAKTSERSOPRG 60

Db 1 MSTLPKPKRKTKRNTNRPTDKFPGGQIVGVYLPRRGRLGRVTRAKTSERSOPRG 60

Qy 61 RRQPIPKARQPOGRHWAQPGYWPMLYNEGGWAGWNLSPRSRSPWGPDTRRSRNLG 120

Db 61 RRQPIPKARQPOGRHWAQPGYWPMLYNEGGWAGWNLSPRSRSPWGPDTRRSRNLG 120

Qy 121 KVIDTUTCGFADLGMYIPWVGAPLGGAVALAHLGVRAIEDGINYATGNLPGCSFSFLA 180

Db 121 KVIDTUTCGFADLGMYIPWVGAPLGGAVALAHLGVRAIEDGINYATGNLPGCSFSFLA 180

Qy 181 LNSCLTPASA 191

Db 181 LNSCLTPASA 191

RESULT 21

S21471

Db 61 RQPIPKARRPEGRAGAWAQPGYWPMLYNEGGWAGWNLSPRSRSPWGPDTRRSRNLG 120

Qy 116-191/Prd: capsid protein; envelope protein

C;Species: hepatitis C virus

C;Date: 20-Feb-1995 #sequence_revision 19-Apr-1996 #text_change 17-Nov-2000

C;Accession: S21471

R;Morgan, W.K.
Submitted to the EMBL Data Library, April 1992

A;Reference number: S21471

A;Accession: S21471

A;Molecule type: genomic RNA

A;Residues: 1-369 <MOG>

A;Cross-references: EMBL:X65946; NID:95946; PIDN:CAA46717.1; PID:959467

C;Superfamily: hepatitis C virus genome polyprotein

C;Keywords: capsid protein; envelope protein; glycoprotein; polyprotein

Query Match Best Local Similarity 90.6%; Score 947; DB 2; Length 369; Matches 173; Conservative 6; Mismatches 12; Indels 0; Gaps 0;

Qy 1 MSTLPKPKRKTKRNTNRPTDKFPGGQIVGVYLPRRGRLGRVTRAKTSERSOPRG 60

Db 1 MSTLPKPKRKTKRNTNRPTDKFPGGQIVGVYLPRRGRLGRVTRAKTSERSOPRG 60

Qy 61 RRQPIPKARQPOGRHWAQPGYWPMLYNEGGWAGWNLSPRSRSPWGPDTRRSRNLG 120

Db 61 RRQPIPKARQPOGRHWAQPGYWPMLYNEGGWAGWNLSPRSRSPWGPDTRRSRNLG 120

Qy 121 KVIDTUTCGFADLGMYIPWVGAPLGGAVALAHLGVRAIEDGINYATGNLPGCSFSFLA 180

Db 121 KVIDTUTCGFADLGMYIPWVGAPLGGAVALAHLGVRAIEDGINYATGNLPGCSFSFLA 180

Qy 181 LNSCLTPASA 191

Db 181 LNSCLTPASA 191

RESULT 22

JQ1926 polyprotein - hepatitis C virus (isolate HCV-476)

N; Contains: C protein; E1 protein; B2/NS1 protein

C; Species: hepatitis C virus

C; Date: 1994 #sequence_revision 03-May-1994 #text_change 17-Nov-2000

C; Accession: JQ1926

R; Abe, K.; Inchauspe, G.; Fujisawa, K.

J. Gen. Virol. 73, 275-279, 1992

A; Title: Genomic characterization and mutation rate of hepatitis C virus isolated from a

A; Reference number: JQ1925; MUID:93019030; PMID:1383400

A; Accession: JQ1926

A; Molecule type: mRNA

A; Residues: 1-533 <A3B>

A; Cross-references: DDBJ:D10687

C; Superfamily: hepatitis C virus genome polyprotein

C; Keywords: polyprotein

F; 1-192383/Product: E1 protein #status predicted <E1R>

F; 1-523/Product: E2/NS1 protein #status predicted <E2P>

Query Match 90.6%; Score 947; DB 2; Length 523;

Best Local Similarity 90.1%; Pred. No. 1.2e-72; Mismatches 7; Indels 12; Gaps 0; Gaps 0;

Matches 172; Conservative 7; Mismatches 12; Indels 0; Gaps 0; Gaps 0;

QY 1 MSTLPKPKRKTKNTNRPTDVKPGGGQIVGAVYLPLRRGPRGLVRAITRSERSOPRG 60

Db 1 MSTLPKPKRKTKNTNRPTDVKPGGGQIVGAVYLPLRRGPRGLVRAITRSERSOPRG 60

QY 61 RQPIPKARQPOCRHAWQPGYWPMLYNEGCWAGWILSPRSRPHGPNDRRSNG 120

Db 61 RQPIPKARQPOCRHAWQPGYWPMLYNEGCWAGWILSPRSRPHGPNDRRSNG 120

QY 121 KVIDTLCGFADLIMGYIPVVGAPLGGAVALAHGVRAILEDGINYATGNLPGCSFSIHLIA 180

Db 121 KVIDTLCGFADLIMGYIPVVGAPLGGAVALAHGVRAILEDGINYATGNLPGCSFSIHLIA 180

QY 181 LLSCLTTPASA 191

Db 181 LLSCLTTPASA 191

RESULT 23

S18032 genome polyprotein - hepatitis C virus (isolate JK4) (fragment)

N; Contains: core protein; envelope protein 1; nonstructural protein 2; NS1/E2 protein

C; Species: hepatitis C virus

C; Variety: isolate JK4

C; Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 17-Nov-2000

C; Session: S18032

C; R; Honma, M.; Kaneko, S.; Masashi, U.; Kobayashi, K.; Murakami, S.

submitted to the EMBL Data Library, September 1991

A; Description: Sequence analysis of putative structural regions of Hepatitis C Virus iso

A; Reference number: S18029

A; Accession: S18032

A; Molecule type: genomic RNA

A; Residues: 1-782 <HON>

A; Cross-references: EMBL:611594

A; Experimental source: isolate JK4

C; Superfamily: hepatitis C virus genome polyprotein

C; Keywords: capsid protein; core protein; envelope protein; glycoprotein; nonstructural

F; 1-192383/Product: core protein #status predicted <MAT1>

F; 1-192383/Product: envelope protein 1 #status predicted <MAT2>

F; 384-733/Product: NS1/E2 protein #status predicted <MAT3>

F; 734-782/Product: nonstructural protein 2 (fragment) #status predicted <MAT4>

Query Match 90.6%; Score 947; DB 2; Length 782;

Best Local Similarity 91.6%; Pred. No. 1.7e-72; Mismatches 4; Indels 0; Gaps 0;

Matches 174; Conservative 4; Mismatches 12; Indels 0; Gaps 0;

QY 1 MSTLPKPKRKTKNTNRPTDVKPGGGQIVGAVYLPLRRGPRGLVRAITRSERSOPRG 60

Db 1 MSTLPKPKRKTKNTNRPTDVKPGGGQIVGAVYLPLRRGPRGLVRAITRSERSOPRG 60

RESULT 24

S41288 genome polyprotein - hepatitis C virus (fragment)

N; Contains: core protein; envelope protein; NS1 protein

C; Species: hepatitis C virus

C; Date: 06-Jan-1995 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000

C; Accession: S41288

R; Seelig, R.

Submitted to the EMBL Data Library, December 1993

A; Reference number: S41288

A; Accession: S41288

A; Molecule type: genomic RNA

A; Residues: 1-492 <SER>

A; Cross-references: EMBL:X76918

C; Superfamily: hepatitis C virus genome polyprotein

C; Keywords: capsid protein; core protein; envelope protein; nonstructural protein; polypro

F; 1-192372/Product: envelope protein #status predicted <ENV>

F; 373-492/Product: NS1 protein (fragment) #status predicted <NS1>

Query Match 90.5%; Score 946; DB 2; Length 492;

Best Local Similarity 90.1%; Pred. No. 1.3e-72; Mismatches 9; Indels 10; Gaps 0; Gaps 0;

Matches 172; Conservative 9; Mismatches 10; Indels 0; Gaps 0; Gaps 0;

QY 1 MSTLPKPKRKTKNTNRPTDVKPGGGQIVGAVYLPLRRGPRGLVRAITRSERSOPRG 60

Db 1 MSTLPKPKRKTKNTNRPTDVKPGGGQIVGAVYLPLRRGPRGLVRAITRSERSOPRG 60

QY 61 RQPIPKARQPOCRHAWQPGYWPMLYNEGCWAGWILSPRSRPHGPNDRRSNG 120

Db 61 RQPIPKARQPOCRHAWQPGYWPMLYNEGCWAGWILSPRSRPHGPNDRRSNG 120

QY 121 KVIDTLCGFADLIMGYIPVVGAPLGGAVALAHGVRAILEDGINYATGNLPGCSFSIHLIA 180

Db 121 KVIDTLCGFADLIMGYIPVVGAPLGGAVALAHGVRAILEDGINYATGNLPGCSFSIHLIA 180

QY 181 LLSCLTTPASA 191

Db 181 LLSCLTTPASA 191

RESULT 25

PC2061 genome polyprotein N2 - hepatitis C virus

N; Contains: envelope protein E1; nonstructural protein B2/NS1

C; Species: hepatitis C virus

C; Date: 14-Jul-1994 #sequence_revision 14-Jul-1994 #text_change 17-Nov-2000

C; Accession: PC2061

C; Li, J.S.; Vitvitski, L.I.; Tong, S.P.; Trepo, C.

Biochem. Biophys. Res. Commun. 199, 1474-1481, 1994

A; Title: Identification of the third major genotype of hepatitis C virus in France.

A; Reference number: PC2060; MUID:9419774; PMID:8147893

A; Accession: PC2061

A; Molecule type: mRNA

A; Residues: 1-411 <LTJ>

A; Cross-references: GB:Lu2355; NID:9410169; PID:AAA20155.1; PID:9410170

C; Superfamily: hepatitis C virus genome polyprotein

C; Keywords: ATP; capsid protein; envelope protein; glycoprotein; nonstructural protein; I

F; 192-383/Product: envelope protein E1 #status predicted <SPB>

F;384-411;Product: nonstructural protein B2/NS1 #status predicted <NP>
 F;196,209,234,305,325/Binding site: carbohydrate (Asn) (covalent) #status predicted
 Query Match 90.3%; Score 944; DB 2; Length 411;
 Best Local Similarity 90.1%; Pred. No. 1; 7e-72; 8; Mismatches 11; Indels 0; Gaps 0;
 Matches 172; Conservative 8;
 C;Species: hepatitis C virus
 C;Accession: A44150
 C;Date: 27-Jun-1994 #sequence_revision 27-Jun-1994 #text_change 17-Nov-2000
 C;Accession: A44150
 C;Species: hepatitis C virus (isolate HC-J6)
 R;Ching, W.M.; Wyrchowski, C.; Beach, M.J.; Wang, H.; Davies, C.L.; Carl, M.; Bradley, D.
 Proc. Natl. Acad. Sci. U.S.A. 89, 3190-3194, 1992
 A;Title: Interaction of immune sera with synthetic peptides corresponding to the structure

Qy 1 MSTLPKQORKTKENTNRPTDVKPGGQIVGVYVILPRRGRLGVRATKTSERSOPRG 60
 Db 1 MSTLPKQORKTKENTNRPTDVKPGGQIVGVYVILPRRGRLGVRATKTSERSOPRG 60
 Qy 61 RRQPIPKARQPOGRHWAQPGYWPWLYNGEGCWAGWILLSPRSRSPHGPNDPDRRSNLG 120
 Db 61 RRQPIPKARQPOGRHWAQPGYWPWLYNGEGCWAGWILLSPRSRSPHGPNDPDRRSNLG 120
 Qy 121 KVIDTLCGFAIDLGYIPVVGAPLGCVAAALAHGVRAIEDGINYATGNLPGCSFSIFLA 180
 Db 121 KVIDTLCGFAIDLGYIPVVGAPLGCVAAALAHGVRAIEDGINYATGNLPGCSFSIFLA 180
 181 LLSCLTPASA 191
 181 LPSCLIHPAS 191

RESULT 26
 JQ0883 genome polyprotein - hepatitis C virus (strain J7) (fragments)
 N;Contains: NS5 protein
 C;Species: hepatitis C virus
 C;Accession: JQ0883
 C;Description: The 5'-terminal and 3'-terminal sequences of the genomic RNA of hepatitis C; Accession: JQ0883
 R;Okamoto, H.
 submitted to JIPID, January 1991
 A;Description: The 5'-terminal and 3'-terminal sequences of the genomic RNA of hepatitis C; Accession: JQ0879
 A;Residues: 1-874 <OXA>
 A;Experimental source: strain J7
 C;Superfamily: hepatitis C virus genome polyprotein
 C;Keywords: polyprotein
 F;510-874/Product: NS5 protein (fragment) #status predicted <NS5>

Query Match 90.0%; Score 941; DB 2; Length 874;
 Best Local Similarity 89.5%; Pred. No. 6e-72; 6; Mismatches 14; Indels 0; Gaps 0;
 Matches 171; Conservative 6;

Qy 1 MSTLPKQORKTKENTNRPTDVKPGGQIVGVYVILPRRGRLGVRATKTSERSOPRG 60
 Db 1 MSTLPKQORKTKENTNRPTDVKPGGQIVGVYVILPRRGRLGVRATKTSERSOPRG 60
 Qy 61 RRQPIPKARQPOGRHWAQPGYWPWLYNGEGCWAGWILLSPRSRSPHGPNDPDRRSNLG 120
 Db 61 RRQPIPKARQPOGRHWAQPGYWPWLYNGEGCWAGWILLSPRSRSPHGPNDPDRRSNLG 120
 Qy 121 KVIDTLCGFAIDLGYIPVVGAPLGCVAAALAHGVRAIEDGINYATGNLPGCSFSIFLA 180
 Db 121 KVIDTLCGFAIDLGYIPVVGAPLGCVAAALAHGVRAIEDGINYATGNLPGCSFSIFLA 180
 181 LLSCLTPASA 191
 181 LPSCLIHPAS 191

RESULT 27
 A44150 structural protein - hepatitis C virus
 C;Species: hepatitis C virus
 C;Accession: A44150
 C;Date: 27-Jun-1994 #sequence_revision 27-Jun-1994 #text_change 17-Nov-2000
 C;Accession: A44150
 C;Species: hepatitis C virus (isolate HC-J6)
 N;Contains: capid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5)

RESULT 28
 JQ0881 genome polyprotein - hepatitis C virus (strain J6) (fragments)
 N;Contains: NS5 protein
 C;Species: hepatitis C virus
 C;Accession: JQ0881
 C;Description: The 5'-terminal and 3'-terminal sequences of the genomic RNA of hepatitis C; Accession: JQ0881
 R;Okamoto, H.
 submitted to JIPID, January 1991
 A;Residues: 1-874 <OXA>
 A;Reference number: JQ0879
 A;Accession: JQ0881
 A;Molecule type: genomic RNA
 A;Residues: 1-874 <OXA>
 A;Experimental source: strain J6
 C;Superfamily: hepatitis C virus genome polyprotein
 C;Keywords: polyprotein
 F;510-874/Product: NS5 protein (fragment) #status predicted <NS5>

Query Match 89.9%; Score 939; DB 2; Length 874;
 Best Local Similarity 89.9%; Pred. No. 8.9e-72; 8; Mismatches 13; Indels 0; Gaps 0;
 Matches 170; Conservative 8;

Qy 1 MSTLPKQORKTKENTNRPTDVKPGGQIVGVYVILPRRGRLGVRATKTSERSOPRG 60
 Db 1 MSTLPKQORKTKENTNRPTDVKPGGQIVGVYVILPRRGRLGVRATKTSERSOPRG 60
 Qy 61 RRQPIPKARQPOGRHWAQPGYWPWLYNGEGCWAGWILLSPRSRSPHGPNDPDRRSNLG 120
 Db 61 RRQPIPKARQPOGRHWAQPGYWPWLYNGEGCWAGWILLSPRSRSPHGPNDPDRRSNLG 120
 Qy 121 KVIDTLCGFAIDLGYIPVVGAPLGCVAAALAHGVRAIEDGINYATGNLPGCSFSIFLA 180
 Db 121 KVIDTLCGFAIDLGYIPVVGAPLGCVAAALAHGVRAIEDGINYATGNLPGCSFSIFLA 180
 181 LLSCLTPASA 191
 181 LPSCLIHPAS 191

RESULT 29
 JQ1303 genome polyprotein - hepatitis C virus (isolate HC-J6)
 N;Contains: capid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5)

Page 11

C;Species: hepatitis C virus
C;Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 17-Nov-2000
C;Accession: JQ1303
R;Okamoto, H.; Okada, S.; Sugiyama, Y.; Kurai, K.; Iizuka, H.; Machida, A.; Miyakawa, Y.
J;Gen. Virol. 72, 2697-2704, 1991
A;Title: Nucleotide sequence of the genomic RNA of hepatitis C virus isolated from a human
A;Reference number: JQ1303; MUID: 92044490; PMID: 1658196
A;Accession: JQ103
A;Molecule type: genomic RNA
A;Residues: 1-3033 <OKA>
A;Cross-references: GB:D00941; NID:g221650; PIDN:BAA00792.1; PID:g221651
A;Experimental source: isolate HC-J6 from a Japanese individual
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: ATP; glycoprotein; hydrolase; P-loop; polyprotein; serine proteinase; transm
F;2-115/Product: capsid protein C #status predicted <CPC>
F;116-191/Product: envelope protein M #status predicted <EPM>
F;192-389/Product: major envelope protein E #status predicted <MEB>
F;390-733/Product: nonstructural protein NS1 #status predicted <NS1>
F;734-1010/Product: nonstructural protein NS2 #status predicted <NS2>
F;1-1619/Product: hepacivirin #status predicted <NS3>
F;1-1321/Region: nucleic-acid-binding motif B
F;1-1323/Region: DEKH motif
F;1620-1866/Product: nonstructural protein NS4a #status predicted <NA4>
F;1867-2017/Product: nonstructural protein NS4b #status predicted <NA4B>
F;2018-33033/Product: nonstructural protein NS5 #status predicted <N05>
F;196-209, 234, 305, 325, 417, 423, 430, 448-477, 534, 542, 558, 578, 627, 649, 1091, 1217, 1259, 2038, 28

R; Kato, N.; Ootsuyama, Y.; Okoshi, S.; Nakawa, T.; Mori, S.; Hijikata, M.; Shimotohno, K.; Biochem. Biophys. Res. Commun. 181, 279-285, 1991
A; Title: Distribution of plaural VCV types in Japan.
A; Reference number: PQ0554; PMID:92068204; PMID:1720309
A; Accession: P00554

A;Accession: JQ1303
 A;Reference number:
 A;Molecule type: genomic RNA
 A;Residues: 1-3033 <OKA>
 A;Cross-references: GB:D00944; NID:g221650; PIDN:BA00792_1; PID:g221651
 A;Experimental source: isolate HC-J6 from a Japanese individual
 C;Superfamily: hepatitis C virus genome polyprotein
 C;Keywords: ATP; glycoprotein; hydrolase; p-loop; hydrolytic; protein; serine proteinase; trans
 F;2-15/Product: capsid protein C #status predicted <CPC>
 F;116-191/Product: envelope protein M #status predicted <EPM>
 F;192-389/Product: major envelope protein E #status predicted <MEB>
 F;390-733/Product: nonstructural protein NS1 #status predicted <NS1>
 F;394-1010/Product: nonstructural protein NS2 #status predicted <NS2>
 F;1-1619/Product: hepacivirin #status predicted <NS3>
 F;1-1321/Region: nucleotide-binding motif B
 F;1-1323/Region: DEKH motif
 F;1620-1866/Product: nonstructural protein NS4A #status predicted <NA4>
 F;1867-2017/Product: nonstructural protein NS4B #status predicted <NB4>
 F;187-2013/Region: nonstructural protein NS5 #status predicted <NS5>
 F;196,209,233,299,305,417,423,430,448,477,534,542,558,578,627,649,1091,1217,1259,2038,28
 Query Match 89.9%; Score 939; DB 1; Length 3033;
 Best Local Similarity 89.0%; Pred. No. 2.9e-71;
 Matches 170; Conservative 8; Mismatches 13; Indels 0; Gaps 0;
 Qy 1 MSTLPKPKRTKRNTRPDKPGGSGIVGGYLLPRGRPLGVRAKTRKTSRSQRG 60
 Db 1 MSTLPKPKRTKRNTRPDKPGGSGIVGGYLLPRGRPLGVRAKTRKTSRSQRG 60
 Qy 61 RROPIPKARQPOGRHMAQPGYPWMLLYGNGCGWAGWLSPRGSRPHWGNDPRRSRUG 120
 Db 61 RROPIPKDRSRTGKSWKGPKPGWPWLYGNEGCGWAGWLSPRGSRPTWGPDPRHSRLG 120
 Qy 121 KVDTLTCGFADLMGYIPVVGAPLGGAVALAHGVRALEDGINATGNILPGCSFSIFLA 180
 Db 121 KVDTLTCGFADLMGYIPVVGAPLGGAVALAHGVRALEDGINATGNILPGCSFSIFLA 180
 Qy 121 KVDTLTCGFADLMGYIPVVGAPLGGAVALAHGVRALEDGINATGNILPGCSFSIFLA 180
 Db 121 KVDTLTCGFADLMGYIPVVGAPLGGAVALAHGVRALEDGINATGNILPGCSFSIFLA 180
 Qy 181 LLSCTVTPASA 191
 Db 181 LLSCTVTPASA 191

38

NC_001608 genome polyprotein - hepatitis C virus (strain HC-J8) Nt-Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5 Protein NS4a; nonstructural protein NS4b; nonstructural protein NS5 C-Species: hepatitis C virus C-Date: 31-Dec-1992 #sequence revision 31-Dec-1992 #text_change 19-Jan-2001 C-Accession: A4250; PQ0397; PQ0559 R.Okamoto, H.; Kuroki, K.; Okada, S.I.; Yamamoto, K.; Iizuka, H.; Tanaka, T.; Fukuda, S.; Virology 188, 331-341, 1992 A-Title: Full-length sequence of a hepatitis C virus genome having poor homology to reported hepatitis C viruses A-Reference number: A4250; NUID:9223032; PMID:1314459 A-Accession: A4250 A-Molecule type: genomic RNA A-Residues: 1-3033 <OKA> A-Cross-references: GB:D10988; GB:D01221; NID:9221608; PIDN:BAA01761; PID:9221609 P-Chai, S.W.; Mcomish, R.; Holmes, R.C.; Dow, R.; Peutherer, J.F.; Rollert, R.; Yan, P.I.

A:Cross-references: GB:D1098; GB:D01221; NID:9221608; PID:9221609
A:Cross-references: GB:D1098; GB:D01221; NID:9221608; PID:9221609
A:Chan, S.W.; McOmisch, H.F.; Holmes, E.C.; Dow, B.; Peutherford, J.F.; Follett, E.; Yap, P.I.
J. Gen. Virol. 73, 1131-1141, 1992.
A:Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship to e
A:Reference number: PQ0397
A:Accession: PQ0397
A:Molecule type: genomic RNA
A:Residues: 2678-2754 <CHA>
A:Cross-references: DDB:D10134
A:Experimental source: isolate E-b12

This Page Blank (uspto)

(EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein NS5B (P66) (RNA-directed RNA polymerase) (EC 2.7.7.48).]

OS Hepatitis C virus (isolate H) (HCV).

OC Viruses; serNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Hepacivirus;

OX NCBI_TAXID=11108;

RN [1]

RP SQUENCE FROM N.A.

RX MEDLINE=92052256; PubMed=1658800;

RX Inchauspe G., Zabedee S., Lee D.H.H., Sugitani M., Nasoff M., Prince A.M.;

RT "Genomic structure of the human prototype strain H of hepatitis C virus: comparison with American and Japanese isolates.";

RJ Proc. Natl. Acad. Sci. U.S.A. 88:10292-10296(1991).

RN [2]

RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1207-1657.

RX MEDLINE=97331322; PubMed=1498764;

RX YAO N., Neeson T., Cable M., Hong Z., Kwong A.D., Le H.V., Weber P.C.; Kim J.L., Morgenstern K.A., Griffith J.P., Dwyer M.D., Thomson J.A., Murcko M.A., Lin C., Caron P.R.;

RT Structure of the hepatitis C virus RNA helicase domain.;

RJ Nat. Struct. Biol. 4:463-467(1997).

RN [3]

RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1192-1657.

RX MEDLINE=98154321; PubMed=1493270;

RX Kim J.L., Morgenstern K.A., Griffith J.P., Dwyer M.D., Thomson J.A.,

RA Murcko M.A., Lin C., Caron P.R.;

RT Oligonucleotide: the crystal structure provides insights into the mode of unwinding.;

RL Structure 6:89-100(1998).

CC -- FUNCTION: PROTEASE NS2 IS RESPONSIBLE FOR THE CLEAVAGE OF NS2-NS3.

CC -- FUNCTION: PROTEASE NS3 IS RESPONSIBLE FOR THE CLEAVAGE OF NS3-NS4A, NS4A-NS4B, NS4B-NS5A AND NS4A-NS5B.

CC -- FUNCTION: NS4A FORMS A COMPLEX WITH NS3 AND IS ESSENTIAL FOR THE ACTIVATION OF NS3.

CC -- FUNCTION: NS4A SEEMS TO HAVE A TRANSCRIPTIONAL ACTIVATORY ROLE.

CC -- FUNCTION: NS4B IS A RNA-DEPENDENT RNA POLYMERASE THAT PLAYS AN ESSENTIAL ROLE IN THE VIRUS REPLICATION.

CC -- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral precursor polyproteins, commonly with Asp or Glu in the P5 position, Cys or Thr in P1 and Ser or Ala in P1'.

CC -- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate + {RNA} (N).

CC -- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: E1 AND E2. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MNA.

CC -- PTM: THE STRUCTURAL PROTEINS C, E1 AND E2 ARE PRODUCED BY PROTEOLYTIC PROCESSING BY THE HOST SIGNAL PEPTIDASES.

CC -- SIMILARITY: THE NS2 PROTEASE BELONGS TO PEPTIDASE FAMILY 829.

CC -- SIMILARITY: THE NS3 PROTEASE BELONGS TO PEPTIDASE FAMILY 829.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; M67463; AA44534.1; -.

DR PIR: A36814; GNNWCH.

DR PDB: 1HEI; 25-NOV-98.

DR PDB: 1ALV; 16-FEB-99.

DR PDB: 1A1R; 17-JUN-98.

DR MEROPS: S29_001; -.

DR MEROPS: U39_001; -.

DR TRANSFAC: T04155; -.

DR InterPro; IPR009003; Cys_Ser_trypsin.

DR InterPro; IPR001410; DEAD.

DR InterPro; IPR002522; HCV_capsid.

DR InterPro; IPR002521; HCV_core.

DR InterPro; IPR002319; HCV_env.

DR InterPro; IPR002531; HCV_NS1.

FT DR InterPro; IPR002519; HCV_NS2.

FT DR InterPro; IPR00745; HCV_NS4a.

FT DR InterPro; IPR001490; HCV_NS4b.

FT DR InterPro; IPR002868; HCV_NS5a.

FT DR InterPro; IPR003166; Helicase_C.

FT DR InterPro; IPR004109; Peptidase_C29.

FT DR InterPro; IPR007095; RNA_Dol DS_PS.

FT DR InterPro; IPR007094; RNA_poli_PSVir.

FT DR Pfam; PF01538; HCV_NS2; 1.

FT DR Pfam; PF02307; HCV_NS3; 1.

FT DR Pfam; PF01006; HCV_NS4a; 1.

FT DR Pfam; PF01001; HCV_NS4b; 1.

FT DR Pfam; PF01506; HCV_NS5a; 1.

FT DR Pfam; PF00371; helicase_C; 1.

FT DR Pfam; PF00598; Viral_RdRp; 1.

FT DR SMART; SM00487; DEXD_C.

FT DR SMART; SM00622; HCV_NS1; 1.

KW Core protein; Glycoprotein; Transferase; RNA-directed RNA polymerase; Polyprotein; Glycoprotein; Envelope protein; Helicase; ATP-binding; Core protein; Coat protein; Transmembrane; Nonstructural protein; Hydrolase; Serine protease; 3D-structure.

FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE CHAIN 1 191 383 CAPSID PROTEIN C.

FT CHAIN 192 384 746 ENVELOPE GLYCOPROTEIN E2.

FT CHAIN 747 809 PROTEIN P7.

FT CHAIN 810 1026 NONSTRUCTURAL PROTEIN NS2.

FT CHAIN 1027 1657 PROTEASE/HELICASE NS3.

FT CHAIN 1658 1711 NONSTRUCTURAL PROTEIN NS4A.

FT CHAIN 1712 1972 NONSTRUCTURAL PROTEIN NS4B.

FT CHAIN 1973 2420 NONSTRUCTURAL PROTEIN NSSA.

FT CHAIN 2421 3011 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT TRANSMEM 347 369 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT NP_BIND 1230 1237 ATP (POTENTIAL).

FT SITE 1316 1319 DECH BOX.

FT CARBOHD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHD 476 476 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHD 540 540 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT STRAND 1224 1226 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT TURN 1232 1233

FT TURN 1239 1238

FT HELIX 1239 1246

FT TURN 1247 1248

FT STRAND 1251 1255

FT HELIX 1258 1271

FT TURN 1272 1272

FT STRAND 1277 1278

FT TURN 1281 1282

FT STRAND 1283 1285

FT STRAND 1291 1293

FT HELIX 1296 1301

FT STRAND 1302 1303

PT STRAND 1312 1316
 PT TURN 1317 1319
 PT HELIX 1323 1335
 PT TURN 1336 1340
 PT STRAND 1343 1347
 PT TURN 1352 1353
 PT TURN 1360 1361
 PT STRAND 1362 1366
 PT STRAND 1368 1369
 PT STRAND 1373 1375
 PT TURN 1376 1377
 PT STRAND 1378 1380
 PT HELIX 1382 1385
 PT STRAND 1389 1393
 PT HELIX 1397 1409
 PT TURN 1410 1411
 PT STRAND 1414 1417
 PT TURN 1419 1420
 PT STRAND 1432 1436
 PT TURN 1438 1439
 PT STRAND 1450 1453
 PT STRAND 1456 1463
 PT STRAND 1471 1478
 PT STRAND 1480 1488
 PT HELIX 1481 1488
 PT TURN 1489 1490
 PT STRAND 1497 1501
 PT STRAND 1507 1507
 PT STRAND 1511 1511
 PT HELIX 1514 1527
 PT HELIX 1532 1544
 PT STRAND 1550 1550
 PT HELIX 1555 1564
 PT HELIX 1570 1578
 PT TURN 1579 1580
 PT HELIX 1584 1597
 PT TURN 1598 1598
 PT HELIX 1606 1611
 PT TURN 1614 1618
 PT STRAND 1622 1623
 PT STRAND 1627 1627
 PT STRAND 1635 1636
 PT HELIX 1640 1652
 SQ SEQUENCE 3011 AA; 327142 MW: 772CB29CCD94753 CRC64;

Query Match 93.4%; Score 976; DB 1; Length 3011;
 Best Local Similarity 92.7%; Pred. No. 1.4e-74; Caches 177; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

DE Genome polyprotein [Contains: Capsid protein C (core protein) (P22); Envelope glycoprotein E1 (GP32); (GP35); Envelope glycoprotein E2 (GP68); (GP70); (NS1); Protein P7; Nonstructural protein NS2 (P21)]
 DE (EC 3.4.22.-); Protease/Helicase NS3 (PRO) [Hepacivirin]
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein NS5B (P66); (P70) [RNA-directed RNA polymerase] (EC 2.7.7.48)].
 DE NS5B (P66); Nonstructural protein NS5A (P56); Nonstructural protein NS5B (P66); (P70) [RNA-directed RNA polymerase] (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate 1) (HCV).
 OC Viruses, ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxId=11104;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91172836; PubMed=1848704;
 RA Choo Q.-L., Richman K.H., Berger K., Lee C., Dong C.,
 RA Gallegos C., Coit D., Medina-Selby A., Barr P.J., Weiner A.J.,
 RA Bradley D.W., Kuo G., Houghton M.;
 RT "Genetic organization and diversity of the hepatitis C virus.";
 RL Proc. Natl. Acad. Sci. U.S.A. 88:2451-2455 (1991).
 CC FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are hydrophobic, suggesting a possible membrane-related function. NS3 and NS5 may play a role in the viral RNA replication.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral precursor polyprotein, commonly with Asp or Glu in the P6 position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate + [RNA] (N).
 CC -!- SUBUNIT: The virion of this virus is a nucleocapsid covered by a lipoprotein envelope. The envelope consists of two proteins: protein M and glycoprotein E. The nucleocapsid is a complex of protein C and mRNA.
 CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL: M62221; AAA45876.1; -.
 DR PIR: A39166; GMWC3.
 DR PDB: 1AVV; 1E-BEB-99.
 DR PDB; IHET; 25-NOV-98.
 MEROPS; S29.001; -.
 MEROPS; U39.001; -.
 DR InterPro: IPR009003; CYB_Ser_trypsin.
 DR InterPro; IPR01410; DEAD.
 DR InterPro; IPR02522; HCV capsid.
 DR InterPro; IPR02521; HCV core.
 DR InterPro; IPR02525; HCV_env.
 DR InterPro; IPR02531; HCV_NS1.
 DR InterPro; IPR02518; HCV_NS2.
 DR InterPro; IPR00745; HCV_NS4a.
 DR InterPro; IPR01490; HCV_NS4b.
 DR InterPro; IPR02868; HCV_NS5a.
 DR InterPro; IPR02166; HCV_RdRP.
 DR InterPro; IPR01650; Helicase_C.
 DR InterPro; IPR004109; Peptidase_C29.
 DR InterPro; IPR007095; RNA_poli_Ds_Ps.
 DR InterPro; IPR007094; RNA_poli_PsVir.
 DR Pfam; PF01543; HCV_capsid.
 DR Pfam; PF01542; HCV_core.
 DR Pfam; PF01539; HCV_env.
 DR Pfam; PF01560; HCV_NS1.
 DR Pfam; PF01538; HCV_NS2.
 DR Pfam; PF02907; HCV_NS3.
 DR Pfam; PF01006; HCV_NS4a.
 DR Pfam; PF01001; HCV_NS4b.
 DR Pfam; PF01505; HCV_NS5a.
 DR Pfam; PF02071; helicase_C.
 DR Pfam; PF00998; viral_RdRP.
 DR Pfam; PF00998; viral_RdRP; 1.

RESULT 2
 POLG_HCV1 STANDARD; PRT; 3011 AA.
 AC P26654;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)

| | | | |
|---------------|--|------------|---|
| DR | prodOm: P0186062; HCV_NS1; 1. | DE | Genome polyprotein [Contains: Capsid protein C (Core protein) (P22); |
| DR | SMART; SMO487; DEXDC; 1. | DE | Envelope glycoprotein E1 (GP32); Envelope glycoprotein E2 |
| KW | Protein; Glycoprotein; Transferase; RNA-directed RNA polymerase; | DE | (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21) |
| KW | Core protein; Coat protein; Envelope protein; Helicase; ATP-binding; | DE | (EC 3.4.22.-); Protease/helicase NS3 (P70) [Hepacivirin] |
| KW | Transmembrane; Nonstructural protein; Hydrolase; Serine protease; | DE | [EC 3.4.21.98]; Nonstructural protein NS4A (P4); Nonstructural protein NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein NS5B (P56) (RNA-directed RNA polymerase) [EC 2.7.7.48]. |
| KW | 3D-structure. | DE | Hepatitis C virus (isolate BK) (HCV). |
| INIT_MET | 1 | OS | Virus; |
| FT | 1 | OC | ssRNA positive-strand viruses, no DNA stage; Flaviviridae; |
| FT | 1 | OC | Hepadviridae; |
| FT | 1 | OX | Hepatitis C virus (isolate BK) (HCV). |
| FT | 1 | NCBI_TAXID | 11105; |
| FT | 1 | RN | [1] |
| FT | 1 | RN | SEQUENCE FROM N. A. |
| FT | 1 | RX | MEDLINE=9114698; PubMed=1847440; |
| FT | 1 | RX | Takamizawa A., Mori C., Fukie T., Manabe S., Murakami S., Fujita J., |
| FT | 1 | RX | Onishi E., Andoh T., Yoshida I., Ohayama H.; |
| FT | 1 | RX | "Structure and organization of the hepatitis C virus genome isolated from human carriers"; |
| FT | 1 | RX | J. Virol. 65:1105-1113 (1991). |
| FT | 1 | RX | [2] |
| FT | 1 | RX | SEQUENCE OF 1487-1500. |
| FT | 1 | RX | MEDLINE=9623524; PubMed=8647104; |
| FT | 1 | RX | Borowski P., Heiland M., Oehlmann K., Becker B., Korneteky L.; |
| FT | 1 | RX | "Non-structural protein kinase inactivation mediated by camp-dependent protein kinase."; |
| FT | 1 | RX | Eur. J. Biochem. 237:611-618 (1996). |
| FT | 1 | RX | [3] |
| FT | 1 | RX | X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 1027-1215. |
| FT | 1 | RX | MEDLINE=97015088; PubMed=881916; |
| FT | 1 | RX | Love R.A., Parげe H.E., Wackersham J.A., Hostomsky Z., Habuka N., |
| FT | 1 | RX | Moomaw E.W., Adachi T., Hostomská Z.; |
| FT | 1 | RX | "The crystal structure of hepatitis C virus NS3 protease reveals a trypsin-like fold and a structural zinc binding site."; |
| FT | 1 | RX | Cell 87:331-342 (1996). |
| FT | 1 | RX | [4] |
| FT | 1 | RX | X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1027-1210 AND 1678-1691. |
| FT | 1 | RX | MEDLINE=98227846; PubMed=9568991; |
| FT | 1 | RX | Yan Y., Li Y., Munshi S., Sardana V., Cole J.L., Sardana M., |
| FT | 1 | RX | Steinkuehler C., Tomei L., de Francesco R., Kuo L.C., Cheen Z.; |
| FT | 1 | RX | "Complex of NS3 protease and NS4A peptide of BK strain hepatitis C virus: a 2.2-A resolution structure in a hexagonal crystal form."; |
| FT | 1 | RX | Protein Sci. 7:1877-1887 (1998). |
| FT | 1 | CC | -I- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are hydrophobic, suggesting a possible membrane-related function. NS3 and NS5 may play a role in the viral RNA replication. |
| FT | 1 | CC | -I- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral precursor polyprotein, commonly with Asp or Glu in the P6 position, Cys or Thr in P1 and Ser or Ala in P1. |
| FT | 1 | CC | -I- CATALYTIC ACTIVITY: N nucleoside triphosphate + (RNA) (N). |
| FT | 1 | CC | -I- SUBUNIT: The virion of this virus is a nucleocapsid covered by a lipoprotein envelope. The envelope consists of two proteins: protein M and glycoprotein E. The nucleocapsid is a complex of protein C and mRNAs. |
| FT | 1 | CC | -I- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29. |
| Db | 1 | CC | This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use. By non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@ish-sib.ch). |
| Db | 1 | CC | EMBL: M58335; AA072945.1; -. |
| Db | 1 | CC | PIR; A38465; GNNTTC. |
| Db | 1 | CC | PDB; 1A1Q; 25-MAR-98. |
| Db | 1 | CC | PDB; 1JXP; 14-JAN-98. |
| Db | 1 | CC | PDB; 1NS3; 08-APR-98. |
| Db | 1 | CC | PDB; 1C2P; 15-NOV-00. |
| Db | 1 | CC | PDB; 1CSJ; 08-NOV-99. |
| Db | 1 | CC | PDB; 1CX5; 09-APR-02. |
| Db | 1 | CC | PDB; 1CX6; 10-APR-02. |
| RESULT | 3 | DR | |
| POLG_HCVBK | STANDARD; | PRT; | 3010 AA. |
| ID_POLG_HCVBK | | | |
| AC | P26663; | | |
| DT | 01-AUG-1992 (Rel. 23, Created) | | |
| DT | 01-AUG-1992 (Rel. 23, Last sequence update) | | |
| OY | 181 LISCLTPASA 191 | | |
| Db | 181 LISCLTPASA 191 | | |

| | | | | | |
|----------|---|-----------------|---|-----------------------|--|
| DR | PDB; 1QUV; 26-JUN-00 | FT | STRAND | 1050 | 1050 |
| DR | PDB; BOHM; 20-APR-99. | FT | STRAND | 1059 | 1063 |
| DR | MEROPS; S29_001; - | FT | STRAND | 1068 | 1074 |
| DR | Interpro; IPR003003; Cys_Ser_trypsin. | FT | TURN | 1075 | 1076 |
| DR | Interpro; IPR00322; HCV_capsid. | FT | STRAND | 1077 | 1081 |
| DR | Interpro; IPR00321; HCV_core. | FT | HELIX | 1082 | 1085 |
| DR | Interpro; IPR00251; HCV_env. | FT | TURN | 1086 | 1087 |
| DR | Interpro; IPR00251; HCV_NS1. | FT | STRAND | 1090 | 1092 |
| DR | Interpro; IPR002518; HCV_NS2. | FT | STRAND | 1095 | 1097 |
| DR | Interpro; IPR002518; HCV_NS4a. | FT | STRAND | 1101 | 1103 |
| DR | Interpro; IPR0149; HCV_NS4b. | FT | STRAND | 1104 | 1107 |
| DR | Interpro; IPR002868; HCV_N5a. | FT | STRAND | 1108 | 1112 |
| DR | Interpro; IPR002166; HCV_RdRp. | FT | STRAND | 1120 | 1120 |
| DR | Interpro; IPR0410; Peptidase_C29. | FT | STRAND | 1122 | 1122 |
| DR | Interpro; IPR007095; RNA_Pol_DS_PS. | FT | STRAND | 1129 | 1133 |
| DR | Interpro; IPR007094; RNA_Pol_Pepr. | FT | TURN | 1135 | 1136 |
| PFam | PF01543; HCV_capsid; 1. | FT | STRAND | 1139 | 1144 |
| PFam | PF01542; HCV_core; 1. | FT | STRAND | 1149 | 1157 |
| PFam | PF01539; HCV_env; 1. | FT | HELIX | 1158 | 1161 |
| PFam | PF01560; HCV_NS1; 1. | FT | TURN | 1162 | 1163 |
| PFam | PF01538; HCV_NS2; 1. | FT | TURN | 1165 | 1166 |
| PFam | PF01537; HCV_NS3; 1. | FT | TURN | 1168 | 1171 |
| PFam | PF0106; HCV_NS4a; 1. | FT | TURN | 1172 | 1174 |
| PFam | PF01001; HCV_NS4b; 1. | FT | STRAND | 1175 | 1186 |
| PFam | PF01506; HCV_N5a; 1. | FT | TURN | 1187 | 1188 |
| PFam | PF00398; Viral_RdRp; 1. | FT | STRAND | 1189 | 1197 |
| ProDom | PDB8602; HCV_NS1; 1. | FT | HELIX | 1198 | 1202 |
| DR | SMART; SM00487; DEXD_C. | FT | TURN | 1203 | 1204 |
| KW | Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase; | FT | STRAND | 1204 | 1688 |
| KW | Core protein; Coat protein; Envelope protein; Helicase; Serine protease; | FT | SEQUENCE | 3010 AA; | 327189 MW; |
| KW | Transmembrane; Nonstructural protein; Hydrolase; Serine protease; 3D structure. | FT | SEQUENCE | 3010 AA; | F8422D5ECCFDPD9C CRC64; |
| INIT_MET | 1 | 1 | REMOVED FROM CAPSID PROTEIN C BY THE | 92.3% | Score 965; DB 1; Length 3010; |
| FT | CHAIN | 1 | CAPSID PROTEIN C (POTENTIAL). | Best Local Similarity | Pred. No. 1..2e-73; |
| FT | CHAIN | 116 | MAJOR MATRIX PROTEIN (POTENTIAL). | Matches 176; | Conservative 5; Mismatches 10; Indels 0; Gaps 0; |
| FT | CHAIN | 192 | MAJOR ENVELOPE PROTEIN B (POTENTIAL). | OY | 1 MSTLPKPKRKTKNTMRPRTDVKFEGGQIVGGAVVLRPGCPRTGRATRCKTSENQPRG 60 |
| FT | CHAIN | 384 | NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL). | Db | 1 MSTNPKPKRKTKNTMRPRTDVKFEGGQIVGGAVVLRPGCPRTGRATRCKTSENQPRG 60 |
| FT | CHAIN | 730 | NONSTRUCTURAL PROTEIN NS2 (POTENTIAL). | OY | 61 RRQPPIKARQPOQRGRHQAQPGYFWPMPLWNGEGWAGWGLLSPRSRSPRPWGNPDRRSRLNIG 120 |
| FT | CHAIN | 1615 | PROTEASE/HELLICASE NS3 (POTENTIAL). | Db | 61 RRQPPIKARQPOQRGRHQAQPGYFWPMPLWNGEGWAGWGLLSPRSRSPRPWGNPDRRSRLNIG 120 |
| FT | CHAIN | 1616 | NONSTRUCTURAL PROTEIN NS4A (POTENTIAL). | OY | 121 KVDTIWTGFAIDLGMGYIPVVRGAFLGVAEDCINYATGNLPCCSFSIFLLA 180 |
| FT | CHAIN | 2013 | NONSTRUCTURAL PROTEIN NS4B (POTENTIAL). | Db | 121 KVDTIWTGFAIDLGMGYIPVVRGAFLGVAEDCINYATGNLPCCSFSIFLLA 180 |
| FT | TRANSMEM | 347 | RNA-DIRECTED RNA POLYMERASE (POTENTIAL). | OY | 181 LLSCLTPASA 191 |
| FT | ACT SITE | 1083 | POTENTIAL. | Db | 181 LLSCLTPASA 191 |
| FT | ACT SITE | 1107 | CHARGE RELAY SYSTEM. | OY | 181 LLSCLTPASA 191 |
| FT | ACT SITE | 1165 | CHARGE RELAY SYSTEM. | Db | 181 LLSCLTPASA 191 |
| FT | NP BIND | 1230 | ATP (POTENTIAL). | OY | 181 LLSCLTPASA 191 |
| FT | SITE | 1317 | DICH_BOK. | Db | 181 LLSCLTPASA 191 |
| FT | CARBODY | 1316 | (POTENTIAL). | OY | 181 LLSCLTPASA 191 |
| FT | CARBODY | 196 | (POTENTIAL). | Db | 181 LLSCLTPASA 191 |
| FT | CARBODY | 209 | (POTENTIAL). | OY | 181 LLSCLTPASA 191 |
| FT | CARBODY | 234 | (POTENTIAL). | Db | 181 LLSCLTPASA 191 |
| FT | CARBODY | 250 | (POTENTIAL). | OY | 181 LLSCLTPASA 191 |
| FT | CARBODY | 305 | (POTENTIAL). | Db | 181 LLSCLTPASA 191 |
| FT | CARBODY | 417 | (POTENTIAL). | OY | 181 LLSCLTPASA 191 |
| FT | CARBODY | 423 | (POTENTIAL). | Db | 181 LLSCLTPASA 191 |
| FT | CARBODY | 430 | (POTENTIAL). | OY | 181 LLSCLTPASA 191 |
| FT | CARBODY | 448 | (POTENTIAL). | Db | 181 LLSCLTPASA 191 |
| FT | CARBODY | 532 | (POTENTIAL). | OY | 181 LLSCLTPASA 191 |
| FT | CARBODY | 540 | (POTENTIAL). | Db | 181 LLSCLTPASA 191 |
| FT | CARBODY | 556 | (POTENTIAL). | OY | 181 LLSCLTPASA 191 |
| FT | CARBODY | 576 | (POTENTIAL). | Db | 181 LLSCLTPASA 191 |
| FT | CARBODY | 623 | (POTENTIAL). | OY | 181 LLSCLTPASA 191 |
| FT | CARBODY | 645 | (POTENTIAL). | Db | 181 LLSCLTPASA 191 |
| FT | CARBODY | 2041 | (POTENTIAL). | OY | 181 LLSCLTPASA 191 |
| FT | CARBODY | 2077 | (POTENTIAL). | Db | 181 LLSCLTPASA 191 |
| FT | CARBODY | 2240 | (POTENTIAL). | OY | 181 LLSCLTPASA 191 |
| FT | CARBODY | 2529 | (POTENTIAL). | Db | 181 LLSCLTPASA 191 |
| FT | CARBODY | 2788 | (POTENTIAL). | OY | 181 LLSCLTPASA 191 |
| FT | STRAND | 1031 | (POTENTIAL). | Db | 181 LLSCLTPASA 191 |
| FT | HELIX | 1039 | (POTENTIAL). | OY | 181 LLSCLTPASA 191 |
| FT | HELIX | 1047 | (POTENTIAL). | Db | 181 LLSCLTPASA 191 |
| RESULT 4 | POLG_HCVJA | ID - POLG_HCVJA | STANDARD: | PRT; | 3010 AA. |
| ID | POLG_HCVJA | AC | | | |
| DT | P26662; | DT | 01-AUG-1992 (Rel. 23, Created) | | |
| DT | | DT | 01-AUG-1992 (Rel. 23, Last sequence update) | | |
| DT | | DT | 28-FEB-2003 (Rel. 41, Last annotation update) | | |
| DB | | DB | Genome polyprotein [Contains: Capsid protein C (Core protein) (P22); Envelope glycoprotein E1 (GP32); Envelope glycoprotein E2 (GP68); Protein p7; Nonstructural protein NS2 (P21)] | | |
| DB | | DB | (BC 3..4..22,-); Protease/helicase NS3 (P70) (Hepacivirus) | | |
| DB | | DB | (BC 3..4..21..98); Nonstructural protein NS4A (P4); Nonstructural protein NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein NS5B (P66) (RNA-directed RNA polymerase) (BC 2..7..7..48)]. | | |
| DB | | DB | HSSB (P66) (P70) (RNA-directed RNA polymerase) (BC 2..7..7..48)]. | | |
| OS | | OS | Hepatitis C virus (Isolate Japanese) (HCV). | | |
| OC | | OC | Virus; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; | | |
| OC | | OC | Hepadviridae; | | |
| RN | | RN | NCBI_TaxID-1116; SEQUENCE FROM N.A.. | [1] | |

RX MEDLINE=91088550; PubMed=2175903;
 RA Kato N., Hijikata M., Ootsuyama Y., Nakagawa M., Ohkoshi S.,
 RA Sugimura T., Shimotohno K.;
 RT "Molecular cloning of the human hepatitis C virus genome from
 Japanese patients with non-A, non-B hepatitis";
 RT Proc. Natl. Acad. Sci. U.S.A. 87:9524-9528(1990).
 RL [2]
 RP DISCUSSION OF SEQUENCE.
 RX MEDLINE=91192160; PubMed=1849488;
 RA Okoshi S., Shimotohno K.;
 RA Okoshi N., Hijikata M., Nakagawa M., Ootsuyama Y., Muraiso K.,
 RA "Molecular structure of the Japanese hepatitis C viral genome.";
 RT FEBS Lett. 280:325-328(1991).
 CC -!- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are
 hydrophobic, suggesting a possible membrane-related function. NS3
 and NS5 may play a role in the viral RNA replication.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 precursor polyprotein, commonly with Asp or Glu in the P6
 position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -!- SUBUNIT: The virion of this virus is a nucleocapsid covered by a
 lipoprotein envelope. The envelope consists of two proteins:
 protein M and glycoprotein E. The nucleocapsid is a complex of
 protein C and mRNA.
 CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial/
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

DR EMBL; D90208; BA14233.1; -.

DR PIR; A39253; GNWYCJ.

DR HSSP; P26653; 1XRP.

DR MEROPS; S29_001; -.

DR InterPro; IPR00903; Cys_Ser_trypsin.

DR InterPro; IPR00410; DEAD.

DR InterPro; IPR002522; HCV_capsid.

DR InterPro; IPR002521; HCV_core.

DR InterPro; IPR002519; HCV_env.

DR InterPro; IPR002531; HCV_NS1.

DR InterPro; IPR002518; HCV_NS2.

DR InterPro; IPR000745; HCV_NS4.

DR InterPro; IPR001490; HCV_NS4b.

DR InterPro; IPR002868; HCV_NS5a.

DR InterPro; IPR002166; HCV_RdRp.

DR InterPro; IPR001650; Helicase_C.

DR InterPro; IPR004109; Peptidase_C29.

DR InterPro; IPR007095; RNA_pol_Ds_ps.

DR InterPro; IPR000794; RNA_pol_Psvr.

DR Pfam; PF0143; HCV_capsid; 1.

DR Pfam; PF0142; HCV_core; 1.

DR Pfam; PF0139; HCV_env; 1.

DR Pfam; PF0160; HCV_NS1; 1.

DR Pfam; PF0138; HCV_NS2; 1.

DR Pfam; PF02907; HCV_NS3; 1.

DR Pfam; PF0106; HCV_NS4a; 1.

DR Pfam; PF0101; HCV_NS4b; 1.

DR Pfam; PF0106; HCV_NS5a; 1.

DR Pfam; PF0071; helicase_C; 1.

DR Pfam; PF0098; viral_RdRp; 1.

DR PRODOM; PD186062; HCV_NS1; 1.

DR SMART; SM0487; DEXD_C; 1.

DR Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;

KW Core protein; Coat protein; Envelope protein; Helicase; Peptidase;

KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease.

FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE

FT SEQENCE FROM N_A.

RX MEDLINE=92230232; PubMed=1314459;

| Query | Match | Score | Length |
|-----------------------|---|---------------------|--------|
| Best Local Similarity | 92.3% | 965 | 3010; |
| Matches | 176; | Pred. No. 1. 2e-73; | |
| Conservative | 5; | Mismatches | 10; |
| | | Indels | 0; |
| | | Gaps | 0; |
| QY | 1 MSTLPLKPRQKTKNTNRRPTDKFPGCGQIVGVYLPRRGRPLGYTRATRKTSERSQPRG | 60 | |
| Db | 1 M\$TNPKCQRKTKNTNRRPTDKFPGCGQIVGVYLPRRGRPLGYTRATRKTSERSQPRG | 60 | |
| QY | 61 RQQPIPKARQPOGRHMAQPGYPPWLYNGEGWAGMLSPRSRPSRPHGPNPDRRSRNLG | 120 | |
| Db | 61 RQQPIPKARQPEGRTPAQPGYPWPWLYNGEGWAGMLSPRSRPSRPHGPNPDRRSRNLG | 120 | |
| QY | 121 KIDTLTGCFADLMGYIPVWVGAPLGGVAAALHGVRALDGINATCNLPCSFSLILLA | 180 | |
| Db | 121 KIDTLTGCFADLMGYIPVWVGAPLGGVAAALHGVRALDGINATCNLPCSFSLILLA | 180 | |
| QY | 181 LIISCLTPASA 191 | | |
| Db | 181 LIISCLTPASA 191 | | |

RESULT 5

POIIG HCVJ2 ID POIIG HCVJ2 STANDARD; PRT; 513 AA.

AC P27959; DT 01-AUG-1992 (Rel. 23, Last sequence update)

DT 01-AUG-1992 (Rel. 23, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22); Envelope glycoprotein E1 (GP32) (Gp35); Envelope glycoprotein E2 (GP68) (GP70) (NS1)] (Fragment).

DE Hepatitis C virus (isolate HC-J2) (HCV).

GS Hepatitis C virus (isolate HC-J2) (HCV).

OC Viruses; SERNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Hepacivirus.

OK NCBI TAXID:1111;

RN [1]

RA Okamoto H.; Kurai K.; Okada S.I.; Yamamoto K.; Lizuka H.,
 RA Tanaka T.; Fukuda S.; Tsuda F.; Mishiro S.;
 RT "Full-length sequence of a hepatitis C virus genome having poor
 homology to reported isolates: comparative study of four distinct
 RT genotypes";
 RL Viralology 188:331-341(1992).
 -I- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are
 hydrophobic, suggesting a possible membrane-related function. NS3
 and NS5 may play a role in the viral RNA replication.
 -I- SUBUNIT: The virion of this virus is a nucleocapsid covered by a
 lipoprotein envelope. The envelope consists of two proteins:
 protein M and glycoprotein E. The nucleocapsid is a complex of
 protein C and mRNA.

This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

CC DR EMBL; D10074; BAA00968.1; .-.

CC DR InterPro; IPR002522; HCV_capsid.

CC DR InterPro; IPR002519; HCV_env.

CC DR InterPro; IPR002531; HCV_NS1.

CC DR Pfam; PF01543; HCV_capsid; 1.

CC DR Pfam; PF01442; HCV_core; 1.

CC DR Pfam; PF01539; HCV_env; 1.

DR Pfam; PF01550; HCV_NS1; 1.

DR Prodom; PDB86062; HCV_NS1.

KW Transmembrane; Nonstructural protein; Coat protein; Envelope protein;

FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 FT CHAIN 1 115 CELLULAR AMINOPEPTIDASE.
 FT CHAIN 116 191 CASID PROTEIN C (POTENTIAL).
 FT CHAIN 192 383 MATRIX PROTEIN (POTENTIAL).
 FT CHAIN 384 >513 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
 FT TRANSMEM 347 369 NONSTRUCTURAL PROTEIN NS1 (POTENTIAL).
 FT CARBOHYD 196 196 POTENTIAL.
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 233 233 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 234 250 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 250 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 417 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 430 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 448 513 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 513 AA; 555704 MW; 943F31B3514CDEF3 CRC64;

Query Match 91.8%; Score 959; DB 1; Length 513;
 Best Local Similarity 91.6%; Pred. No. 6; Jee-74;
 Matches 175; Conservative 5; Mismatch 11; Indels 0; Gaps 0;

OY 1 MSTLPKPORTKTKNTNRNRPDKVPGGGQIVGGYLURRGPPGLVGRATRKRSERSQPRG 60
 Db 1 MSTLPKPKRKTAKNTNRNRPDKVPGGGQIVGGYLURRGPPGLVGRATRKRSERSQPRG 60
 OY 61 ROPPIPKARQPOQRHWAQPGPYNPWLYGNBGNAGWLLSPRSRPHMGPNDPRRSRNG 120
 OY 61 ROPPIPKARPKRPEGRAWAQPGPYNPWLYGNBGNAGWAGWLSPRSRPSWGPDTRRRSRNLG 120
 OY 121 KVIDLTGCGPAIDLMGYIPLVGAPLGAARALAHGVRLIEDSVNATGNLPGCSFSIILIA 180
 Db 121 KVIDLTGCGPAIDLMGYIPLVGAPLGAARALAHGVRLIEDSVNATGNLPGCSFSIILIA 180
 OY 181 LLSCLTPASA 191
 Db 181 LLSCLTPASA 191

RESULT 6
 ID POIG_HCVTW
 ID POIG_HVWTW
 STANDARD; PRT; 3010 AA.
 AC P29846;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP65); Envelope glycoprotein E2
 DE (GP68) (GP70); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA Polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate Taiwan) (HCV).
 OC Viruses; RNA positive-strand viruses, no DNA stage; Flaviviridae;
 Hepacivirus.
 OK NCBI_TaxID=31645;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92230206; PubMed=13114419;
 RA Chen P.J.; Lin M.H.; Tai K.F.; Liu P.C.; Lin C.J.J.; Chen D.S.;
 RT "The Taiwanese hepatitis C virus genome: sequence determination and
 mapping the 5' termini of viral genomic and antigenomic RNA.;"
 CC Virology 188:102-113(1992).
 CC -I- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are
 hydrophobic, suggesting a possible membrane-related function. NS3
 and NS5 may play a role in the viral RNA replication.
 CC -I- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 precursor polyprotein, commonly with Asp or Glu in the P6
 position. Cys or Thr in P1 and Ser or Ala in P1'.
 CC -I- CATALYTIC ACTIVITY: N nucleoside triphosphate +
 CC {RNA} (N).
 CC -I- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 DR MEROPS; S29_001; .-.
 DR MEROPS; U39_001; .-.
 DR InterPro; IPR005003; Cys_Ser_trypsin.
 DR InterPro; IPR01410; DEAD.
 DR InterPro; IPR002522; HCV_capsid.
 DR InterPro; IPR002521; HCV_core.
 DR InterPro; IPR002519; HCV_env.
 DR InterPro; IPR002531; HCV_NS1.
 DR InterPro; IPR002518; HCV_NS2.
 DR InterPro; IPR000745; HCV_NS4.
 DR InterPro; IPR01490; HCV_NS4b.
 DR InterPro; IPR02868; HCV_NS5a.
 DR InterPro; IPR002166; HCV_RdRp.
 DR InterPro; IPR01650; Helicase_C.
 DR InterPro; IPR04109; Peptidase_C29.
 DR InterPro; IPR007095; RNA_Polymerase.
 DR InterPro; IPR007094; RNA_Polymerase.
 DR Pfam; PF01543; HCV_capsid; 1.
 DR Pfam; PF01542; HCV_core; 1.

| RESULT | 8 | POI | HCVUT | STANDARD: | PRT; | 3010 AA. |
|--------|---------|-----|---|-----------|---|----------|
| A | 000269; | DR | 01-APR-1993 (Rel. 25, Created) | DT | 10-OCT-2003 (Rel. 42, Last annotation update) | DE |
| D | | DR | Genome polyprotein [Contains: Capsid protein C (Core protein) (P22); Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2 (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21) (EC 3.4.22.-); Protease/heilicase NS3 (P70) (Hepacivirus) (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein NS5B (P66); RNA-directed RNA polymerase (EC 2.7.7.48)]. | DE | | |
| B | 181 | Db | LMSCLTPASA 191 | Qy | 121 KVIDLTGFGFDLMLGYTIVGAGLGGVAAALHGRYALEDGNYATGMLPGSFSIILA 180 | DR |
| C | | Qy | 121 KVIDLTGFGFDLMLGYTIVGAGLGGVAAALHGRYALEDGNYATGMLPGSFSIILA 180 | Db | 121 KVIDLTGFGFDLMLGYTIVGAGLGGVAAALHGRYALEDGNYATGMLPGSFSIILA 180 | DR |
| D | | Db | LMSCLTPASA 191 | Qy | 181 LMSCLTPASA 191 | DR |
| E | | | | POI | POI | POI |
| F | | | | POI | POI | POI |
| G | | | | POI | POI | POI |
| H | | | | POI | POI | POI |
| I | | | | POI | POI | POI |
| J | | | | POI | POI | POI |
| K | | | | POI | POI | POI |
| L | | | | POI | POI | POI |
| M | | | | POI | POI | POI |
| N | | | | POI | POI | POI |
| O | | | | POI | POI | POI |
| P | | | | POI | POI | POI |
| Q | | | | POI | POI | POI |
| R | | | | POI | POI | POI |
| S | | | | POI | POI | POI |
| T | | | | POI | POI | POI |
| U | | | | POI | POI | POI |
| V | | | | POI | POI | POI |
| W | | | | POI | POI | POI |
| X | | | | POI | POI | POI |
| Y | | | | POI | POI | POI |
| Z | | | | POI | POI | POI |
| AA | | | | POI | POI | POI |
| BB | | | | POI | POI | POI |
| CC | | | | POI | POI | POI |
| DD | | | | POI | POI | POI |
| EE | | | | POI | POI | POI |
| FF | | | | POI | POI | POI |
| GG | | | | POI | POI | POI |
| HH | | | | POI | POI | POI |
| II | | | | POI | POI | POI |
| JJ | | | | POI | POI | POI |
| KK | | | | POI | POI | POI |
| LL | | | | POI | POI | POI |
| MM | | | | POI | POI | POI |
| NN | | | | POI | POI | POI |
| OO | | | | POI | POI | POI |
| PP | | | | POI | POI | POI |
| QQ | | | | POI | POI | POI |
| RR | | | | POI | POI | POI |
| SS | | | | POI | POI | POI |
| TT | | | | POI | POI | POI |
| UU | | | | POI | POI | POI |
| VV | | | | POI | POI | POI |
| WW | | | | POI | POI | POI |
| XX | | | | POI | POI | POI |
| YY | | | | POI | POI | POI |
| ZZ | | | | POI | POI | POI |
| AA | | | | POI | POI | POI |
| BB | | | | POI | POI | POI |
| CC | | | | POI | POI | POI |
| DD | | | | POI | POI | POI |
| EE | | | | POI | POI | POI |
| FF | | | | POI | POI | POI |
| GG | | | | POI | POI | POI |
| HH | | | | POI | POI | POI |
| II | | | | POI | POI | POI |
| JJ | | | | POI | POI | POI |
| KK | | | | POI | POI | POI |
| LL | | | | POI | POI | POI |
| MM | | | | POI | POI | POI |
| NN | | | | POI | POI | POI |
| OO | | | | POI | POI | POI |
| PP | | | | POI | POI | POI |
| QQ | | | | POI | POI | POI |
| RR | | | | POI | POI | POI |
| SS | | | | POI | POI | POI |
| TT | | | | POI | POI | POI |
| UU | | | | POI | POI | POI |
| VV | | | | POI | POI | POI |
| WW | | | | POI | POI | POI |
| XX | | | | POI | POI | POI |
| YY | | | | POI | POI | POI |
| ZZ | | | | POI | POI | POI |
| AA | | | | POI | POI | POI |
| BB | | | | POI | POI | POI |
| CC | | | | POI | POI | POI |
| DD | | | | POI | POI | POI |
| EE | | | | POI | POI | POI |
| FF | | | | POI | POI | POI |
| GG | | | | POI | POI | POI |
| HH | | | | POI | POI | POI |
| II | | | | POI | POI | POI |
| JJ | | | | POI | POI | POI |
| KK | | | | POI | POI | POI |
| LL | | | | POI | POI | POI |
| MM | | | | POI | POI | POI |
| NN | | | | POI | POI | POI |
| OO | | | | POI | POI | POI |
| PP | | | | POI | POI | POI |
| QQ | | | | POI | POI | POI |
| RR | | | | POI | POI | POI |
| SS | | | | POI | POI | POI |
| TT | | | | POI | POI | POI |
| UU | | | | POI | POI | POI |
| VV | | | | POI | POI | POI |
| WW | | | | POI | POI | POI |
| XX | | | | POI | POI | POI |
| YY | | | | POI | POI | POI |
| ZZ | | | | POI | POI | POI |
| AA | | | | POI | POI | POI |
| BB | | | | POI | POI | POI |
| CC | | | | POI | POI | POI |
| DD | | | | POI | POI | POI |
| EE | | | | POI | POI | POI |
| FF | | | | POI | POI | POI |
| GG | | | | POI | POI | POI |
| HH | | | | POI | POI | POI |
| II | | | | POI | POI | POI |
| JJ | | | | POI | POI | POI |
| KK | | | | POI | POI | POI |
| LL | | | | POI | POI | POI |
| MM | | | | POI | POI | POI |
| NN | | | | POI | POI | POI |
| OO | | | | POI | POI | POI |
| PP | | | | POI | POI | POI |
| QQ | | | | POI | POI | POI |
| RR | | | | POI | POI | POI |
| SS | | | | POI | POI | POI |
| TT | | | | POI | POI | POI |
| UU | | | | POI | POI | POI |
| VV | | | | POI | POI | POI |
| WW | | | | POI | POI | POI |
| XX | | | | POI | POI | POI |
| YY | | | | POI | POI | POI |
| ZZ | | | | POI | POI | POI |
| AA | | | | POI | POI | POI |
| BB | | | | POI | POI | POI |
| CC | | | | POI | POI | POI |
| DD | | | | POI | POI | POI |
| EE | | | | POI | POI | POI |
| FF | | | | POI | POI | POI |
| GG | | | | POI | POI | POI |
| HH | | | | POI | POI | POI |
| II | | | | POI | POI | POI |
| JJ | | | | POI | POI | POI |
| KK | | | | POI | POI | POI |
| LL | | | | POI | POI | POI |
| MM | | | | POI | POI | POI |
| NN | | | | POI | POI | POI |
| OO | | | | POI | POI | POI |
| PP | | | | POI | POI | POI |
| QQ | | | | POI | POI | POI |
| RR | | | | POI | POI | POI |
| SS | | | | POI | POI | POI |
| TT | | | | POI | POI | POI |
| UU | | | | POI | POI | POI |
| VV | | | | POI | POI | POI |
| WW | | | | POI | POI | POI |
| XX | | | | POI | POI | POI |
| YY | | | | POI | POI | POI |
| ZZ | | | | POI | POI | POI |
| AA | | | | POI | POI | POI |
| BB | | | | POI | POI | POI |
| CC | | | | POI | POI | POI |
| DD | | | | POI | POI | POI |
| EE | | | | POI | POI | POI |
| FF | | | | POI | POI | POI |
| GG | | | | POI | POI | POI |
| HH | | | | POI | POI | POI |
| II | | | | POI | POI | POI |
| JJ | | | | POI | POI | POI |
| KK | | | | POI | POI | POI |
| LL | | | | POI | POI | POI |
| MM | | | | POI | POI | POI |
| NN | | | | POI | POI | POI |
| OO | | | | POI | POI | POI |
| PP | | | | POI | POI | POI |
| QQ | | | | POI | POI | POI |
| RR | | | | POI | POI | POI |
| SS | | | | POI | POI | POI |
| TT | | | | POI | POI | POI |
| UU | | | | POI | POI | POI |
| VV | | | | POI | POI | POI |
| WW | | | | POI | POI | POI |
| XX | | | | POI | POI | POI |
| YY | | | | POI | POI | POI |
| ZZ | | | | POI | POI | POI |
| AA | | | | POI | POI | POI |
| BB | | | | POI | POI | POI |
| CC | | | | POI | POI | POI |
| DD | | | | POI | POI | POI |
| EE | | | | POI | POI | POI |
| FF | | | | POI | POI | POI |
| GG | | | | POI | POI | POI |
| HH | | | | POI | POI | POI |
| II | | | | POI | POI | POI |
| JJ | | | | POI | POI | POI |
| KK | | | | POI | POI | POI |
| LL | | | | POI | POI | POI |
| MM | | | | POI | POI | POI |
| NN | | | | POI | POI | POI |
| OO | | | | POI | POI | POI |
| PP | | | | POI | POI | POI |
| QQ | | | | POI | POI | POI |
| RR | | | | POI | POI | POI |
| SS | | | | POI | POI | POI |
| TT | | | | POI | POI | POI |
| UU | | | | POI | POI | POI |
| VV | | | | POI | POI | POI |
| WW | | | | POI | POI | POI |
| XX | | | | POI | POI | POI |
| YY | | | | POI | POI | POI |
| ZZ | | | | POI | POI | POI |
| AA | | | | POI | POI | POI |
| BB | | | | POI | POI | POI |
| CC | | | | POI | POI | POI |
| DD | | | | POI | POI | POI |
| EE | | | | POI | POI | POI |
| FF | | | | POI | POI | POI |
| GG | | | | POI | POI | POI |
| HH | | | | POI | POI | POI |
| II | | | | POI | POI | POI |
| JJ | | | | POI | POI | POI |
| KK | | | | POI | POI | POI |
| LL | | | | POI | POI | POI |
| MM | | | | POI | POI | POI |
| NN | | | | POI | POI | POI |
| OO | | | | POI | POI | POI |
| PP | | | | POI | POI | POI |
| QQ | | | | POI | POI | POI |
| RR | | | | POI | POI | POI |
| SS | | | | POI | POI | POI |
| TT | | | | POI | POI | POI |
| UU | | | | POI | POI | POI |
| VV | | | | POI | POI | POI |
| WW | | | | POI | POI | POI |
| XX | | | | POI | POI | POI |
| YY | | | | POI | POI | POI |
| ZZ | | | | POI | POI | POI |
| AA | | | | POI | POI | POI |
| BB | | | | POI | POI | POI |
| CC | | | | POI | POI | POI |
| DD | | | | POI | POI | POI |
| EE | | | | POI | POI | POI |
| FF | | | | POI | POI | POI |
| GG | | | | POI | POI | POI |
| HH | | | | POI | POI | POI |
| II | | | | POI | POI | POI |
| JJ | | | | POI | POI | POI |
| KK | | | | POI | POI | POI |
| LL | | | | POI | PO | |

| | | | | | | | | |
|------------------|---|--|------|--|--|--|--|--|
| Db | 1 | MSTNPKQRKTKRNTYRQDVKPFGQQIVGGVYVLRGPGTIGVATRKRSERSQPRG | 60 | FT CARBOHYD FT CARBOHYD FT CARBOHYD FT CARBOHYD | 418 424 424 431 449 449 | N-LINKED N-LINKED N-LINKED N-LINKED N-LINKED N-LINKED | (GLCNAC. . .) (GLCNAC. . .) | (POTENTIAL) (POTENTIAL) (POTENTIAL) (POTENTIAL) (POTENTIAL) (POTENTIAL) |
| Oy | 61 | RQOPIKAROPQGRHWAQPGYPPWLYNGEGWAGWILSPRSRPRHGPNDRRSRNLG | 120 | | | | | |
| Db | 61 | RQOPIKARRPEGRAWAQPGYPPWLYNGEGWAGWILSPRSRPSRPNSPDTDRRSRLG | 120 | | | | | |
| Oy | 121 | KVIDTLCGFAIDLGYIPVUGAPLGGAARALAHGVRLVEDGVNYATGNLPGCSFSIPLA | 180 | | | | | |
| Db | 121 | KVIDTLCGFAIDLGYIPVUGAPLGGAARALAHGVRLVEDGVNYATGNLPGCSFSIPLA | 180 | | | | | |
| Oy | 181 | LISCLTIPASA 191 | | | | | | |
| Db | 181 | LISCLTIPASA 191 | | | | | | |
| RESULT 9 | | | | | | | | |
| POLG_HCV4 | | STANDARD; | PRT; | 520 AA. | | | | |
| AC | Q01404; | | | | | | | |
| DR | 001-JUL-1993 (Rel. 26, Created) | | | | | | | |
| DR | 01-JUL-1993 (Rel. 26, Last sequence update) | | | | | | | |
| DR | 16-OCT-2001 (Rel. 40, Last annotation update) | | | | | | | |
| DE | Genome polyprotein [Contains: Capsid Protein C (Core protein) (P22); Envelope glycoprotein E1 (GP32); Envelope glycoprotein E2 (GP68); (GP70); (NS1)] (Fragment) | | | | | | | |
| DE | Hepatitis C virus (isolate HCV-476) (HCV) | | | | | | | |
| OS | Hepatitis C virus (isolate HCV-476) (HCV) | | | | | | | |
| OC | Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus | | | | | | | |
| OX | NCBI_TaxID=31643; | | | | | | | |
| RN | [1] | | | | | | | |
| RP | SEQUENCE FROM N.A. | | | | | | | |
| RX | MEDLINE=93019030; PubMed=1383400; | | | | | | | |
| RA | Abe K., Inchauspe G., Fujisawa K.; | | | | | | | |
| RT | "Genomic characterization and mutation rate of hepatitis C virus isolated from a patient who contracted hepatitis during an epidemic of non-A, non-B hepatitis in Japan."; | | | | | | | |
| RT | J. Gen. Virol. 73:2735-2739(1992). | | | | | | | |
| CC | -!- SUBUNIT: The virion of this virus is a nucleocapsid covered by a lipoprotein envelope. The envelope consists of two proteins: protein M and glycoprotein E. The nucleocapsid is a complex of protein C and mRNA. | | | | | | | |
| CC | or send an email to license@ibb-sib.ch). | | | | | | | |
| CC | This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.ibb-sib.ch/announce/) or send an email to license@ibb-sib.ch). | | | | | | | |
| DR | EMBL; D10688; BAA01530.1; --. | | | | | | | |
| DR | InterPro; IPR002522; HCV_capsid. | | | | | | | |
| DR | InterPro; IPR002521; HCV_core. | | | | | | | |
| DR | InterPro; IPR002519; HCV_env. | | | | | | | |
| DR | InterPro; IPR002531; HCV_NS1. | | | | | | | |
| DR | PF01543; HCV_capsid; 1. | | | | | | | |
| DR | PFam; PF01542; HCV_core; 1. | | | | | | | |
| DR | PFam; PF01539; HCV_env; 1. | | | | | | | |
| DR | PFam; PF01540; HCV_NS1; 1. | | | | | | | |
| DR | ProDom; PDB6062; HCV_NS1; 1. | | | | | | | |
| KW | Polyprotein; Glycoprotein; Coat protein; Envelope protein; | | | | | | | |
| FT | Transmembrane; Nonstructural protein | | | | | | | |
| FT | INIT_MET | | | | | | | |
| FT | REMOVED FROM CAPSID PROTEIN C BY THE CELLULAR AMINOPEPTIDASE. | | | | | | | |
| FT | CHAIN | 1 | 115 | | | | | |
| FT | CHAIN | 116 | 191 | | | | | |
| FT | CHAIN | 192 | 383 | | | | | |
| FT | CHAIN | 384 | >520 | | | | | |
| FT | TRANSMEM | 347 | 369 | | | | | |
| FT | CARBOHYD | 196 | 196 | | | | | |
| FT | CARBOHYD | 209 | 209 | | | | | |
| FT | CARBOHYD | 234 | 234 | | | | | |
| FT | CARBOHYD | 305 | 305 | | | | | |
| RESULT 10 | | | | | | | | |
| POLG_HCV7 | | STANDARD; | PRT; | 737 AA. | | | | |
| AC | P27967; | | | | | | | |
| DR | 01-AUG-1992 (Rel. 23, Created) | | | | | | | |
| DR | 01-AUG-1992 (Rel. 23, Last sequence update) | | | | | | | |
| DR | 16-OCT-2001 (Rel. 40, Last annotation update) | | | | | | | |
| DE | Genome polyprotein [Contains: Capsid protein C (Core protein); Matrix protein (Envelope protein M); Major envelope protein E; Nonstructural protein NS1 and NS2] (Fragment) | | | | | | | |
| DE | Hepatitis C virus (isolate HC-37) (HCV) | | | | | | | |
| OS | Hepatitis C virus (isolate HC-37) (HCV) | | | | | | | |
| OC | Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus | | | | | | | |
| OK | NCBI_TaxID=1114; | | | | | | | |
| RN | [1] | | | | | | | |
| RP | SEQUENCE FROM N.A. | | | | | | | |
| RX | MEDLINE=9223023; PubMed=1314459; | | | | | | | |
| RA | Okamoto H., Kurai K., Okuda S.I., Yamamoto K., Lizuka H., Tanaka T., Fukuda S., Tsuda F., Mihiro S.; | | | | | | | |
| RA | "Full-length sequence of a hepatitis C virus genome having poor homology to reported isolates: comparative study of four distinct genotypes". | | | | | | | |
| RR | Virology 188:331-341(1992). | | | | | | | |
| RR | -!- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are hydrophobic, suggesting a possible membrane-related function. NS3 and NS5 may play a role in the viral RNA replication. | | | | | | | |
| CC | -!- SUBUNIT: The virion of this virus is a nucleocapsid covered by a lipoprotein envelope. The envelope consists of two proteins: protein M and glycoprotein E. The nucleocapsid is a complex of protein C and mRNA. | | | | | | | |
| CC | or send an email to license@ibb-sib.ch). | | | | | | | |
| CC | This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.ibb-sib.ch/announce/) or send an email to license@ibb-sib.ch). | | | | | | | |
| DR | EMBL; D10077; BAA00971.1; --. | | | | | | | |
| DR | InterPro; IPR002522; HCV_capsid. | | | | | | | |
| DR | InterPro; IPR002521; HCV_core. | | | | | | | |
| DR | InterPro; IPR002519; HCV_env. | | | | | | | |
| DR | InterPro; IPR002531; HCV_NS1. | | | | | | | |
| DR | PF01543; HCV_capsid; 1. | | | | | | | |
| DR | PF01542; HCV_core; 1. | | | | | | | |
| DR | PF01539; HCV_env; 1. | | | | | | | |
| DR | PF01540; HCV_NS1; 1. | | | | | | | |
| KW | Polyprotein; Glycoprotein; Coat protein; Envelope protein; | | | | | | | |
| FT | Transmembrane; Nonstructural protein | | | | | | | |
| FT | INIT_MET | 1 | 1 | | | | | |
| FT | REMOVED FROM CAPSID PROTEIN C BY THE CELLULAR AMINOPEPTIDASE. | | | | | | | |
| FT | CHAIN | 1 | 115 | | | | | |
| FT | CHAIN | 116 | 191 | | | | | |
| FT | CHAIN | 192 | 383 | | | | | |
| FT | CHAIN | 384 | >520 | | | | | |
| FT | TRANSMEM | 347 | 369 | | | | | |
| FT | CARBOHYD | 196 | 196 | | | | | |
| FT | CARBOHYD | 209 | 209 | | | | | |
| FT | CARBOHYD | 234 | 234 | | | | | |
| FT | CARBOHYD | 305 | 305 | | | | | |

| | |
|-----------------------|--|
| DR | Pfam; PF01542; HCV_core; 1. |
| DR | Pfam; PF01539; HCV_env; 1. |
| DR | Pfam; PF01540; HCV_NS1; 1. |
| DR | ProDom; PDB86062; HCV_NS1; 1. |
| KW | Transmembrane; Nonstructural protein; Coat protein; Envelope protein; |
| KW | INIT_MET |
| FT | REMOVED FROM CAPSID PROTEIN C BY THE CELLULAR AMINOPEPTIDASE. |
| FT | CHAIN 1 115 CAPSID PROTEIN C (POTENTIAL). |
| FT | CHAIN 116 191 MATRIX PROTEIN (POTENTIAL). |
| FT | CHAIN 192 383 MAJOR ENVELOPE PROTEIN E (POTENTIAL). |
| FT | CHAIN 384 733 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL). |
| FT | >737 369 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL). |
| FT | 347 369 POTENTIAL. |
| FT | TRANSMEM 1 115 CAPSID PROTEIN C (POTENTIAL). |
| FT | CHAIN 116 191 MATRIX PROTEIN (POTENTIAL). |
| FT | CHAIN 192 383 MAJOR ENVELOPE PROTEIN E (POTENTIAL). |
| FT | CHAIN 384 733 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL). |
| FT | 347 369 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL). |
| FT | 347 369 POTENTIAL. |
| FT | CARBOHYD 1 115 CAPSID PROTEIN C (POTENTIAL). |
| FT | CARBOHYD 195 209 MATRIX PROTEIN (POTENTIAL). |
| FT | CARBOHYD 209 233 MAJOR ENVELOPE PROTEIN E (POTENTIAL). |
| FT | CARBOHYD 233 299 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL). |
| FT | CARBOHYD 299 305 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL). |
| FT | CARBOHYD 305 347 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL). |
| FT | CARBOHYD 417 423 N-LINKED (GLCNAC. . .) (POTENTIAL). |
| FT | CARBOHYD 423 430 N-LINKED (GLCNAC. . .) (POTENTIAL). |
| FT | CARBOHYD 430 448 N-LINKED (GLCNAC. . .) (POTENTIAL). |
| FT | CARBOHYD 448 477 N-LINKED (GLCNAC. . .) (POTENTIAL). |
| FT | CARBOHYD 477 534 N-LINKED (GLCNAC. . .) (POTENTIAL). |
| FT | CARBOHYD 534 542 N-LINKED (GLCNAC. . .) (POTENTIAL). |
| FT | CARBOHYD 542 558 N-LINKED (GLCNAC. . .) (POTENTIAL). |
| FT | CARBOHYD 558 578 N-LINKED (GLCNAC. . .) (POTENTIAL). |
| FT | CARBOHYD 578 627 N-LINKED (GLCNAC. . .) (POTENTIAL). |
| FT | CARBOHYD 627 649 N-LINKED (GLCNAC. . .) (POTENTIAL). |
| FT | CARBOHYD 649 737 N-LINKED (GLCNAC. . .) (POTENTIAL). |
| FT | NON_TER 737 737 AA; 81691 MW; 67DFAB11854122F2 CRC64; |
| SQ | SEQUENCE FROM N.A. |
| Query Match | 90.0%; Score 941; DB 1; Length 737; |
| Best Local Similarity | 89.5%; Precl. No. 3e-72; |
| Matches | 171; Conservative 6; Mismatches 14; Indels 0; Gaps 0; |
| Qy | 1 MSTRUPKPORTKRTNTRPPTDVKPGGQIVGGYLLPRRGPRGVRATRKTSERSQPRG 60 |
| Db | 1 MSTRUPKPORTKRTNTRPPTDVKPGGQIVGGYLLPRRGPRGVRATRKTSERSQPRG 60 |
| Qy | 61 ROPIPKQRQPRQRHWRQPGYPPRLYGRGGAGWILSPRGRPHMCPNDRRRSLIG 120 |
| Db | 61 ROPIPKQRSTGKSWKGPKGPWPLYGNBGGWAGWILSPRGSRPTWPTDPHRRSRNG 120 |
| Qy | 121 KVIDLTGCFADLGMVPIVNGVAPLGGAVALANQVATEDGINYATGVLPGSPSFIELA 180 |
| Db | 121 KVIDLTGCFADLGMVPIVNGVAPLGGAVALANQVATEDGINYATGVLPGSPSFIELA 180 |
| Qy | 181 LLSCLTTPASA 191 |
| Db | 181 LLSCVTVFVUSA 191 |
| RESULT 11 | |
| POLG_HCVJ6 | STANDARD; PRT; 3033 AA. |
| ID | POLG_HCVJ6 |
| AC | P26660; |
| DT | 01-AUG-1992 (Rel. 23; Last sequence update) |
| DT | 28-PEB-2003 (Rel. 41; Last annotation update) |
| DE | Genome Polyprotein [Contains]: Capsid protein C (Core protein) (P22); |
| DE | Envelope glycoprotein E1 (GP32); Envelope glycoprotein E2 (GP68) (GP70); Protein NS1; Protein P7; Nonstructural protein NS2 (P21) (EC 3.4.22.-); Protease/heilcase NS3 (P71) (Hepacivirus) |
| DE | (EC 3.4.22.-); Protease/heilcase NS3 (P71) (Hepacivirus) |
| DE | Nonstructural protein NS4A (P4); Nonstructural protein NS4B (P5); Nonstructural protein NS5A (P56); Nonstructural protein NS5B (P66); RNA-directed RNA polymerase (EC 2.7.7.48)]. |
| DE | NSSB (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)]. |
| OS | Hepatitis C virus (isolate HC-J6) (HCV). |
| OC | Viruses; ssRNA positive-strand viruses; no DNA stage; Flaviviridae; |
| OC | Hepadviridae; |
| OX | NCBI_TaxID=1113; |
| [1] | |

| | |
|-----------------------|---|
| DE | Genome polyprotein [Contains: Matrix protein (Envelope protein M); Major envelope protein B; Nonstructural protein NS1] (Fragment). |
| ID | POLG_HCVEL |
| STANDARD; | |
| PRT; | 192 AA. |
| RP | SEQUENCE FROM N.A. |
| RX | MEDLINE=91112009; PubMed=1846505; |
| RA | Weiner A.J., Brauer M.J., Rosenblatt J., Richman K.H., Tung J., Crawford K., Bonino F., Saracco G., Choo Q.-L., Houghton M., Han J.H.; |
| OC | Hepatitis C virus (isolate ECL) (HCV). |
| OC | Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus; |
| OC | Reoviridae; |
| OX | NCBI_TaxID=1110; |
| RN | [1] |
| RN | SEQUENCE FROM N.A. |
| RX | MEDLINE=91112009; PubMed=1846505; |
| RA | Weiner A.J., Brauer M.J., Rosenblatt J., Richman K.H., Tung J., Crawford K., Bonino F., Saracco G., Choo Q.-L., Houghton M., Han J.H.; |
| CC | "Variable and hypervariable domains are found in the regions of HCV corresponding to the flavivirus envelope and NS1 proteins and the pestivirus envelope glycoproteins."; |
| RL | Virology 180:842-848(1991). |
| CC | -1 SUBUNIT: The virion of this virus is a nucleocapsid covered by a lipoprotein envelope. The envelope consists of two proteins: |
| CC | "Variable and hypervariable domains are found in the regions of HCV corresponding to the flavivirus envelope and NS1 proteins and the pestivirus envelope glycoproteins."; |
| CC | This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch). |
| CC | This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch). |
| CC | EMBL; X53131; CAA37291; 1. -. |
| CC | InterPro; IPR002521; HCV core. |
| DR | InterPro; IPR002521; HCV core. |
| DR | InterPro; IPR002519; HCV env. |
| DR | InterPro; IPR002519; HCV NS1. |
| DR | Pfam; PF01542; HCV core; 1. |
| DR | Pfam; PF01539; HCV env; 1. |
| DR | Pfam; PF01560; HCV NS1; 1. |
| DR | Prodom; PD18062; HCV NS1; 1. |
| KW | Polyprotein; Glycoprotein; Coat protein; Envelope protein; Transmembrane; Nonstructural protein. |
| FT | NON TER 1 1 |
| FT | CHAIN <1 76 MATRIX PROTEIN (POTENTIAL). |
| FT | CHAIN 77 267 MAJOR ENVELOPE PROTEIN E (POTENTIAL). |
| FT | CHAIN 268 >321 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL). |
| FT | CARBOHYD 80 80 N-LINKED (GLCNAC. . .) (POTENTIAL). |
| FT | CARBOHYD 93 93 N-LINKED (GLCNAC. . .) (POTENTIAL). |
| FT | CARBOHYD 118 118 N-LINKED (GLCNAC. . .) (POTENTIAL). |
| FT | CARBOHYD 189 189 N-LINKED (GLCNAC. . .) (POTENTIAL). |
| FT | CARBOHYD 301 301 N-LINKED (GLCNAC. . .) (POTENTIAL). |
| FT | CARBOHYD 307 307 N-LINKED (GLCNAC. . .) (POTENTIAL). |
| FT | CARBOHYD 314 314 N-LINKED (GLCNAC. . .) (POTENTIAL). |
| FT | CARBOHYD 321 321 N-LINKED (GLCNAC. . .) (POTENTIAL). |
| SQ | SEQUENCE 321 AA; 34238 MN; 2F5DE79F7CT84C8 CRC64; |
| Query Match | 34.5%; Score 361; DB 1; Length 321; |
| Best Local Similarity | 90.7%; Pred. No. 1.5e-23; |
| Matches | 68; Conservative 3; Mismatches 4; Indels 0; Gaps 0; |
| Query Match | 34.5%; Score 361; DB 1; Length 321; |
| Best Local Similarity | 90.7%; Pred. No. 1.5e-23; |
| Matches | 68; Conservative 3; Mismatches 4; Indels 0; Gaps 0; |
| Qy | 117 RNLGKVITLTGCFADLMGVIPVUGAPLGQVAAALAHGVRAIEGIVNTGNLPGCSFSI 176 |
| Db | 1 RNLGKVITLTGCFADLMGVIPVUGAPLGQVAAALAHGVRAIEGIVNTGNLPGCSFSI 60 |
| Qy | 177 FLLAALSCITTPASA 191 |
| Db | 61 FLLAALSCITTPASA 75 |
| RESULT 15 | |
| POLG_HCVB | |
| ID | POLG_HCVB |
| STANDARD; | |
| PRT; | 321 AA. |
| RP | SEQUENCE FROM N.A. |
| RX | MEDLINE=91112009; PubMed=1846505; |
| RA | Weiner A.J., Brauer M.J., Rosenblatt J., Richman K.H., Tung J., Crawford K., Bonino F., Saracco G., Choo Q.-L., Houghton M., Han J.H.; |
| OC | Hepatitis C virus (isolate ECL) (HCV). |
| OC | Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus; |
| OC | Reoviridae; |
| OX | NCBI_TaxID=1110; |
| RN | [1] |
| RN | SEQUENCE FROM N.A. |
| RX | MEDLINE=91112009; PubMed=1846505; |
| RA | Weiner A.J., Brauer M.J., Rosenblatt J., Richman K.H., Tung J., Crawford K., Bonino F., Saracco G., Choo Q.-L., Houghton M., Han J.H.; |
| CC | "Variable and hypervariable domains are found in the regions of HCV corresponding to the flavivirus envelope and NS1 proteins and the pestivirus envelope glycoproteins."; |
| CC | This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch). |
| CC | EMBL; X53131; CAA37291; 1. -. |
| CC | InterPro; IPR002521; HCV core. |
| DR | InterPro; IPR002519; HCV env. |
| DR | InterPro; IPR002519; HCV NS1. |
| DR | Pfam; PF01542; HCV core; 1. |
| DR | Pfam; PF01539; HCV env; 1. |
| DR | Pfam; PF01560; HCV NS1; 1. |
| DR | Prodom; PD18062; HCV NS1; 1. |
| KW | Polyprotein; Glycoprotein; Coat protein; Envelope protein; Transmembrane; Nonstructural protein. |
| FT | NON TER 1 1 |
| FT | CHAIN <1 76 MATRIX PROTEIN (POTENTIAL). |
| FT | CHAIN 77 267 MAJOR ENVELOPE PROTEIN E (POTENTIAL). |
| FT | CHAIN 268 >321 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL). |
| FT | CARBOHYD 80 80 N-LINKED (GLCNAC. . .) (POTENTIAL). |
| FT | CARBOHYD 93 93 N-LINKED (GLCNAC. . .) (POTENTIAL). |
| FT | CARBOHYD 118 118 N-LINKED (GLCNAC. . .) (POTENTIAL). |
| FT | CARBOHYD 189 189 N-LINKED (GLCNAC. . .) (POTENTIAL). |
| FT | CARBOHYD 301 301 N-LINKED (GLCNAC. . .) (POTENTIAL). |
| FT | CARBOHYD 307 307 N-LINKED (GLCNAC. . .) (POTENTIAL). |
| FT | CARBOHYD 314 314 N-LINKED (GLCNAC. . .) (POTENTIAL). |
| FT | CARBOHYD 321 321 N-LINKED (GLCNAC. . .) (POTENTIAL). |
| SQ | SEQUENCE 321 AA; 34238 MN; 2F5DE79F7CT84C8 CRC64; |
| Query Match | 34.5%; Score 361; DB 1; Length 321; |
| Best Local Similarity | 90.7%; Pred. No. 1.5e-23; |
| Matches | 68; Conservative 3; Mismatches 4; Indels 0; Gaps 0; |
| Qy | 117 RNLGKVITLTGCFADLMGVIPVUGAPLGQVAAALAHGVRAIEGIVNTGNLPGCSFSI 176 |
| Db | 1 RNLGKVITLTGCFADLMGVIPVUGAPLGQVAAALAHGVRAIEGIVNTGNLPGCSFSI 60 |
| Qy | 177 FLLAALSCITTPASA 191 |
| Db | 61 FLLAALSCITTPASA 75 |
| RESULT 16 | |
| POLG_HCVH | |
| ID | POLG_HCVH |
| STANDARD; | |
| PRT; | 321 AA. |
| RP | SEQUENCE FROM N.A. |
| RX | MEDLINE=91112009; PubMed=1846505; |
| RA | Weiner A.J., Brauer M.J., Rosenblatt J., Richman K.H., Tung J., Crawford K., Bonino F., Saracco G., Choo Q.-L., Houghton M., Han J.H.; |
| OC | Hepatitis C virus (isolate ECL) (HCV). |
| OC | Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus; |
| OC | Reoviridae; |
| OX | NCBI_TaxID=1110; |
| RN | [1] |
| RN | SEQUENCE FROM N.A. |
| RX | MEDLINE=91112009; PubMed=1846505; |
| RA | Weiner A.J., Brauer M.J., Rosenblatt J., Richman K.H., Tung J., Crawford K., Bonino F., Saracco G., Choo Q.-L., Houghton M., Han J.H.; |
| CC | "Variable and hypervariable domains are found in the regions of HCV corresponding to the flavivirus envelope and NS1 proteins and the pestivirus envelope glycoproteins."; |
| CC | This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch). |
| CC | EMBL; X53131; CAA37291; 1. -. |
| CC | InterPro; IPR002521; HCV core. |
| DR | InterPro; IPR002519; HCV env. |
| DR | InterPro; IPR002519; HCV NS1. |
| DR | Pfam; PF01542; HCV core; 1. |
| DR | Pfam; PF01539; HCV env; 1. |
| DR | Pfam; PF01560; HCV NS1; 1. |
| DR | Prodom; PD18062; HCV NS1; 1. |
| KW | Polyprotein; Glycoprotein; Coat protein; Envelope protein; Transmembrane; Nonstructural protein. |
| FT | NON TER 1 1 |
| FT | CHAIN <1 76 MATRIX PROTEIN (POTENTIAL). |
| FT | CHAIN 77 267 MAJOR ENVELOPE PROTEIN E (POTENTIAL). |
| FT | CHAIN 268 >321 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL). |
| FT | CARBOHYD 80 80 N-LINKED (GLCNAC. . .) (POTENTIAL). |
| FT | CARBOHYD 93 93 N-LINKED (GLCNAC. . .) (POTENTIAL). |
| FT | CARBOHYD 118 118 N-LINKED (GLCNAC. . .) (POTENTIAL). |
| FT | CARBOHYD 189 189 N-LINKED (GLCNAC. . .) (POTENTIAL). |
| FT | CARBOHYD 301 301 N-LINKED (GLCNAC. . .) (POTENTIAL). |
| FT | CARBOHYD 307 307 N-LINKED (GLCNAC. . .) (POTENTIAL). |
| FT | CARBOHYD 314 314 N-LINKED (GLCNAC. . .) (POTENTIAL). |
| FT | CARBOHYD 321 321 N-LINKED (GLCNAC. . .) (POTENTIAL). |
| SQ | SEQUENCE 321 AA; 34238 MN; 2F5DE79F7CT84C8 CRC64; |
| Query Match | 34.5%; Score 361; DB 1; Length 321; |
| Best Local Similarity | 90.7%; Pred. No. 1.5e-23; |
| Matches | 68; Conservative 3; Mismatches 4; Indels 0; Gaps 0; |
| Qy | 117 RNLGKVITLTGCFADLMGVIPVUGAPLGQVAAALAHGVRAIEGIVNTGNLPGCSFSI 176 |
| Db | 1 RNLGKVITLTGCFADLMGVIPVUGAPLGQVAAALAHGVRAIEGIVNTGNLPGCSFSI 60 |
| Qy | 177 FLLAALSCITTPASA 191 |
| Db | 61 FLLAALSCITTPASA 75 |

DE Genome polyprotein [Contains: Matrix protein (Envelope protein M);
 DE Major envelope protein E; Nonstructural protein NS1] (Fragment).
 OS Hepatitis C virus (isolate TH) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OC NCBI_TaxID=1117;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=9112009; PubMed=1846505;
 RA Weiner A.J., Brauer M.J., Rosenblatt J., Richman K.H., Tung J.,
 RA Crawford K., Bonino F., Saracco G., Choo Q.-L., Houghton M.,
 RA Han J.H.;
 RT "Variable and hypervariable domains are found in the regions of HCV
 corresponding to the flavivirus envelope and NS1 proteins and the
 pestivirus envelope glycoproteins.";
 RL Virology 180:842-848(1991).
 CC --! SUBUNIT: The virion of this virus is a nucleocapsid covered by a
 CC -lipoprotein envelope. The envelope consists of two proteins:
 CC protein M and glycoprotein E. The nucleocapsid is a complex of
 CC protein C and mRNA.
 CC .
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC modified and this statement is not removed. Usage by and for commercial/
 CC or send an email to license@ibb-sib.ch).
 CC .
 DR EMBL; X53134; CAA37244.1; -.
 DR InterPro; IPR002521; HCV_core.
 DR InterPro; IPR002519; HCV_env.
 DR Pfam; PF01542; HCV_NS1.
 DR Pfam; PF01539; HCV_env; 1.
 DR Pfam; PF01560; HCV_NS1; 1.
 DR ProDom; PDB86062; HCV_NS1; 1.
 KW Transmembrane; Nonstructural protein.
 PT NON_TER 1 1
 PT CHAIN <1 1
 PT CHAIN 76 257 MATRIX PROTEIN (POTENTIAL).
 PT CHAIN 268 >321 MATRIX PROTEIN (POTENTIAL).
 PT CARBOHYD 80 80 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
 PT CARBOHYD 93 93 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
 PT CARBOHYD 118 118 N-LINKED (GLCNAC. . .) (POTENTIAL).
 PT CARBOHYD 189 189 N-LINKED (GLCNAC. . .) (POTENTIAL).
 PT CARBOHYD 301 307 N-LINKED (GLCNAC. . .) (POTENTIAL).
 PT CARBOHYD 314 314 N-LINKED (GLCNAC. . .) (POTENTIAL).
 PT NON_TER 321 321 AA; 34074 MW; B2EBB3F521C3B520 CRC64;
 SQ .
 Query Match 34.4%; Score 359; DB 1; Length 321;
 Best Local Similarity 89.3%; Pred. No. 2 2e-23; Matches 67; Conservative 4; Indels 0; Gaps 0;
 Matches 67; Conservative 4; Indels 0; Gaps 0;
 Qy 117 RNLGKVDTLGCFADLMGYIPIVGAIGGVAALAHGVRAEDGINVATGNLPGCSFSI 176
 Db 1 RNLGKVDTLGCFADLMGYIPIVGAIGGVAALAHGVRAEDGINVATGNLPGCSFSI RNLGKVDTLGCFADLMGYIPIVGAIGGVAALAHGVRAEDGINVATGNLPGCSFSI 60
 Qy 177 FLIALLSCLTPASA 191
 Db 61 FLIALLSCLTPASA 75
 RESULT 17
 ID POLG_HCVH7 STANDARD; PRT; 309 AA.
 AC P27955;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=9112009; PubMed=1846505;
 RA Weiner A.J., Brauer M.J., Rosenblatt J., Richman K.H., Tung J.,
 RA Crawford K., Bonino F., Saracco G., Choo Q.-L., Houghton M.,
 RA Han J.H.;
 RT "Variable and hypervariable domains are found in the regions of HCV
 corresponding to the flavivirus envelope and NS1 proteins and the
 pestivirus envelope glycoproteins.";
 RL Virology 180:842-848(1991).
 CC --! SUBUNIT: The virion of this virus is a nucleocapsid covered by a
 CC protein M and glycoprotein E. The nucleocapsid is a complex of
 CC protein C and mRNA.
 CC .
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC modified and this statement is not removed. Usage by and for commercial/
 CC or send an email to license@ibb-sib.ch).
 CC .
 DR EMBL; X53133; CAA37293.1; -.
 DR ProDom; PDB86062; HCV_core.
 DR InterPro; IPR002521; HCV_env.
 DR InterPro; IPR002531; HCV_NS1.
 DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV_env; 1.
 DR Pfam; PF01560; HCV_NS1; 1.
 KW Polyprotein; Glycoprotein; Coat protein; Envelope protein;
 KW Transmembrane; Nonstructural protein.
 PT NON_TER 1 1
 PT CHAIN <1 63 MATRIX PROTEIN (POTENTIAL).
 PT CHAIN 64 255 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
 PT CHAIN 256 >309 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
 PT CARBOHYD 68 68 N-LINKED (GLCNAC. . .) (POTENTIAL).
 PT CARBOHYD 81 81 N-LINKED (GLCNAC. . .) (POTENTIAL).
 PT CARBOHYD 106 106 N-LINKED (GLCNAC. . .) (POTENTIAL).
 PT CARBOHYD 177 177 N-LINKED (GLCNAC. . .) (POTENTIAL).
 PT CARBOHYD 289 289 N-LINKED (GLCNAC. . .) (POTENTIAL).
 PT CARBOHYD 295 295 N-LINKED (GLCNAC. . .) (POTENTIAL).
 PT CARBOHYD 302 302 N-LINKED (GLCNAC. . .) (POTENTIAL).
 PT NON_TER 309 309 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 309 AA; 32922 MW; 6E8-B89C1D0B9E9 CRC64;
 Query Match 28.5%; Score 298; DB 1; Length 309;
 Best Local Similarity 88.9%; Pred. No. 3e-18; Matches 56; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
 Matches 56; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
 Qy 129 GRADLMGYIPIVGAIGGVAALAHGVRAEDGINVATGNLPGCSFSI 188
 Db 1 GRADLMGYIPIVGAIGGVAALAHGVRAEDGINVATGNLPGCSFSI FLIALLSCLTPA 60
 Qy 189 ASA 191
 Db 61 ASA 63
 RESULT 18
 LYXO_CHICK STANDARD; PRT; 420 AA.
 ID LYXO_CHICK STANDARD; PRT; 420 AA.
 AC 005063;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DS Protein-Lysine 6-oxidase precursor (EC 1.4.3.13) (Lysyl oxidase).
 OS Gallus gallus (Chicken).
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;

| OX | NCBI_TaxID=9031; | RESULT 19 |
|--------|--|-----------------|
| RN | [1] SEQUENCE FROM N.A. | PRIORI_CADHI |
| RP | | ID_PRIORI_CADHI |
| RC | | STANDARD; |
| TISSUE | =Embryo; | PRT; |
| RA | MEDLINE=93077497; PubMed=1360009; | 256 AA. |
| RT | "Characterization and developmental expression of chick aortic lysyl-oxidease." | |
| RL | J. Biol. Chem. 267:24199-24206 (1992). | |
| RT | - - FUNCTION: Responsible for the posttranslational oxidative deamination of peptidyl lysine residues in precursors to fibrous collagen and elastin. | |
| CC | - - SUBCELLULAR LOCATION: Extracellular. | |
| CC | - - DEVELOPMENTAL STAGE: Increases between day 8 and 16 of embryonic development, during aortic embryogenesis, in direct proportion to total protein synthesis. | |
| CC | - - PTM: The lysine tyrosylquinone cross-link (LHQ) is generated by condensation of the epsilon-amino group of a lysine with a tyroquinone produced by oxidation of tyrosine. | |
| CC | - - SIMILARITY: Belongs to the Lysyl oxidase family. | |
| CC | This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch). | |
| CC | ----- | |
| CC | DR EMBL; M97881; AAA48942.1; -. | |
| CC | PIR; A45166; A45166; Lysyl oxidase. | |
| CC | DR InterPro; IPR01695; Lysyl oxidase. | |
| CC | DR Pfam; PF01186; Lysyl oxidase_1. | |
| CC | DR PRINTS; PRO0074; LYSTOXIDASE. | |
| CC | DR PRODOM; PDD13887; Lysyl oxidase_1. | |
| CC | DR PROSITE; PS00926; LYSYL_OXIDASE_1. | |
| CC | KW OXIDOREDUCTASE; Copper; Signal; Glycoprotein; HQO. | |
| CC | FT SIGNAL; 1; 21. | |
| CC | FT PROPEP; 2; 153. | |
| CC | FT CHAIN; 154; 420. | |
| CC | FT DOMAIN; 216; 420. | |
| CC | FT DOMAIN; 58; 286. | |
| CC | FT METAL; 295; 295. | |
| CC | FT METAL; 297; 297. | |
| CC | FT METAL; 299; 299. | |
| CC | FT CROSSLINK; 323; 358. | |
| CC | FT MOD_RSS; 358; 358. | |
| CC | FT CARBOHYD; 378; 378. | |
| CC | FT SEQUENCE; 420 AA; 48152 MW; B0CBE81DD625F5C2 CRC64; | |
| CC | Query Match 9.3%; Score 97.5.; DB 1; Length 420; Best Local Similarity 28.2%; Pred. No. 0; 3:8; Matches 35; Conservative 10; Mismatches 30; Indels 49; Gaps 7; CC | |
| CC | CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch). | |
| CC | CC ----- | |
| CC | CC 22 VPKPGGGIVV-----GGYVLLPARG-----PRGCVRATRKTSERSQ 57 | |
| CC | CC 38 IQWENQGQVYSLISQAOYQPARRQGAEPASSPVLLLRGNGSVPRAAAAAARPQPERQ 97 | |
| CC | CC ----- | |
| CC | CC 58 PRGRQRQPKPA--RQPOG-RHWAQPGYWPMLPYGENGCWAGWLISPROS-RPHWGNDPR 113 | |
| CC | CC 98 POKQQPQPRPRSSRRQPLGRHRWFOAGY-----RAPSGSARP---APR 136 | |
| CC | CC ----- | |
| CC | CC 114 RRSR 117 | |
| CC | CC 137 RRPR 140 | |
| CC | CC ----- | |
| CC | CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch). | |
| CC | CC ----- | |
| CC | CC DR EMBL; X31999; CA6363050; 1; -. | |
| CC | CC DR EMBL; S82561; AAD14409; 1; -. | |
| CC | CC DR PIR; S37149; S37149. | |
| CC | CC DR HSPB; P10279; IDWY. | |
| CC | CC DR InterPro; IPR00817; Prion. | |
| CC | CC DR Pfam; PF00377; prion_1. | |
| CC | CC DR Pfam; PF03991; Prion octapep_5. | |
| CC | CC DR PRINTS; PR01341; PRION. | |

DR SMART; SM00157; PRION_1; 1.
 DR PROSITE; PS0001; PRION_1; 1.
 DR PROSITE; PS00706; PRION_2; 1.
 KW Prion; Glycoprotein; GPI-anchor; Repeat; Signal.
 SIGNAL 1 24 BY SIMILARITY.
 FT CHAIN 25 256 MAJOR PRION PROTEIN.
 FT CARBOHYD 184 184 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 200 200 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT DISULFID 182 217 5 X 8 AA TANDEM REPEATS OF P-H-G-G-W-G-
 DOMAIN 54 95 Q. BY SIMILARITY.
 FT REPEAT 54 62 1.
 FT REPEAT 63 70 2.
 FT REPEAT 71 78 3.
 FT REPEAT 79 86 4.
 FT REPEAT 87 95 5.
 FT VARIANT 142 142 I -> M (APPEARS TO BE ASSOCIATED WITH
 DIFFERING DISEASE INCUBATION PERIODS IN
 GOATS EXPERIMENTALLY INFECTED WITH
 ISOLATES OF BOVINE SPONGIFORM
 ENCEPHALOPATHY OR SHEEP SCRAPE).
 SEQUENCE 256 AA; 27897 MW; BDA575R6FD99746 CRC64;
 SQ Query Match 8.9%; Score 93.5; DB 1; Length 256;
 Best Local Similarity 29.2%; Pred. No. 0.5; DB 1;
 Matches 47; Conservative 6; Mismatches 57; Indels 51; Gaps 12;
 FT REPEAT 73 GRHWAQ---GYPWPLYGNEGCGW----AGWLSPRSRPHGPNDPRRSRNKGKV 123
 FT REPEAT 65 GGGWQPHGGWQGP---HGGGWPQPHGGW-GQQGSHSQW--NPKSKPKTNMKHVA 116
 FT REPEAT 124 DLTGCFADIMGYIPIVVGAPLGIVAAALAHGVRAIEDGINY 164
 FT REPEAT 117 GAAAG-----AVVGG-LGGYMLGSAMSRSPLHFQNDY 148
 RESULT 20
 PRIO_FELCA ID PRI0_FELCA STANDARD; PRT; 256 AA.
 AC 018754; 019016; 1.
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 D Major prion protein precursor (PrP).
 D PBP or PrP.
 OS Felis silvestris catus (Cat).
 OC Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.
 OX NCBI_TAXID=9685;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE_Blood, and Brain;
 RA Rohwer R.G.; Edelman J.L.; Protzman J.L.;
 RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 112-235 FROM N.A.
 RA Taylor M.S.; Newton D.J.; Flanagan B.F.; Christmas S.B.;
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: The function of PrP is unknown. PrP is encoded in the host genome and is expressed both in normal and infected cells.
 CC -!- SUBUNIT: PrP has a tendency to aggregate yielding polymers called "rods".
 CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
 CC -!- DISEASE: FOUND IN HIGH QUANTITY IN THE BRAIN OF HUMANS AND ANIMALS INFECTED WITH DEGENERATIVE NEUROLOGICAL DISEASES SUCH AS KURU, CREUTZFEILD-JAKOB DISEASE (CJD), GERSTMANN-STRAUSSLER SYNDROME (GSS), SCRAPPY, BOVINE SPONGIFORM ENCEPHALOPATHY (BSE), TRANSMISSIBLE MINK ENCEPHALOPATHY (TME), ETC.
 CC -!- SIMILARITY: Belongs to the prion family.
 DR -!- CAUTION: Ref.1 sequence seems to be incorrect; it is too close in sequence to that of sheep to be taxonomically correct. We have used Ref.2 sequence in the region where it is available (112-235), but the rest of the sequence probably contains incorrect residues.
 CC -!- DATABASE: NAME-Cat_PRP; NOTE-Web page on cat sequence problems; WWW=http://www.mad-cow.org/cat_prion.html.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 DR EMBL; AR003087; AAB70468; 1; -.
 DR EMBL; 113698; CAR74032; 1; -.
 DR HSSP; P04925; IAG2.
 DR InterPro; IPR00817; CRC64;
 DR SMART; SM00157; PRP; 1.
 DR PROSITE; PS00706; PRION_1; 1.
 DR PROSITE; PS00391; PRION.
 DR Pfam; PF00377; Prion; 1.
 DR Pfam; PF03991; Prion octapep; 5.
 DR PRINTS; PR00341; PRION.
 DR SMART; SM00157; PRP; 1.
 DR PROSITE; PS00706; PRION_2; 1.
 DR PROSITE; PS00391; PRION.
 DR PRION; Glycoprotein; GPI-anchor; Repeat; Signal.
 FT SIGNAL 1 24 MAJOR PRION PROTEIN.
 FT CHAIN 25 256 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 184 184 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT DISULFID 200 217 5 X 8 AA TANDEM REPEATS OF P-H-G-G-W-G-
 DOMAIN 54 95 Q. BY SIMILARITY.
 FT REPEAT 73 GRHWAQ---GYPWPLYGNEGCGW----AGWLSPRSRPHGPNDPRRSRNKGKV 123
 FT REPEAT 65 GGGWQPHGGWQGP---HGGGWPQPHGGW-GQQGSHSQW--NPKSKPKTNMKHVA 116
 FT REPEAT 124 DLTGCFADIMGYIPIVVGAPLGIVAAALAHGVRAIEDGINY 164
 FT REPEAT 117 GAAAG-----AVVGG-LGGYMLGSAMSRSPLHFQNDY 148
 RESULT 21
 PRIO_SHEEP ID PRIO_SHEEP STANDARD; PRT; 256 AA.
 AC P23307;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)
 DT 01-NOV-1991 (Rel. 20, Last annotation update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Major prion protein precursor (PrP).

ON PRRP OR PRP.
 OS Ovis aries (Sheep).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Caprinae; Ovis.
 CC NCBI_TAXID=9940;
 RN [1] SEQUENCE FROM N.A., AND VARIANT GLN-171.
 RP STRAIN=Suffolk; TISSUE=Spleen;
 RX MEDLINE=90207216; PubMed=1969655;
 RA Goldmann W., Hunter N., Foster J.D., Salbaum J.M., Beyreuther K.,
 HOPE J.; "Two alleles of a neural protein gene linked to scrapie in sheep.";
 RT Proc. Natl. Acad. Sci. U.S.A. 87:2476-2480(1990).
 RN [2] SEQUENCE FROM N.A.
 RP TISSUE=Liver;
 RA Inoue S., Watanaabe A., Horiuchi M., Ishiguro N., Shinagawa M.;
 RL Submitted (SEP-1994) to the EMBL/GenBank/DDBJ databases.
 RX [3] SEQUENCE FROM N.A.
 RP STRAIN=Suffolk; TISSUE=Brain;
 RX MEDLINE=95011594; PubMed=1926730;
 RA Westaway D., Zuliani V., Cooper C.M., da Costa M., Neuman S.,
 JENNY A.L., Detwiler L., Prusiner S.B.;
 RT "Homology search for prion protein alleles encoding glutamine-171
 renders sheep susceptible to natural scrapie.";
 RL Genes Dev. 8:959-969(1994).
 RN [4] VARIANTIS SCRAPIE VAL-136; HIS-154 AND GLN-171.
 RP MEDLINE=92013957; PubMed=1681077;
 RA Goldmann W., Hunter N., Benson G., Foster J.D., Hope J.;
 RT "Different scrapie-associated fibril proteins (PrP) are encoded by
 lines of sheep selected for different alleles of the SIP gene.";
 RL J. Gen. Virol. 72:2411-2417(1991).
 RN [5] VARIANTIS SCRAPIE THR-112; VAL-136 AND HIS-154.
 RX MEDLINE=93162675; PubMed=8094373;
 RA Laplanche J.L., Chatzilain J., Westaway D., Thomas S., Dussaucy M.,
 RA Brugere-Picoux J., Launay J.-M.;
 RT "PrP polymorphisms associated with natural scrapie discovered by
 denaturing gradient gel electrophoresis.";
 RL Genomics 15:30-37(1993).
 RN [6] VARIANTIS SCRAPIE VAL-136 AND HIS-171, AND VARIANT HIS-154.
 RX MEDLINE=95205072; PubMed=789744;
 RA Belt P.B.G.M., Mulleman I.H., Schreuder B.E.C., Bos-De Ruijter J.,
 RA Gielkens A.L.J., Smits M.A.;
 RT "Identification of five allelic variants of the sheep PrP gene and
 their association with natural scrapie.";
 RL J. Gen. Virol. 76:509-517(1995).
 RN [7] VARIANTS THR-137; PHE-141 AND GLN-211.
 RX MEDLINE=97042305; PubMed=8878505;
 RA Bossers A., Schreuder B.E.C., Mulleman I.H., Belt P.B.G.M.,
 RA Smits M.A.;
 RT "PrP genotype contributes to determining survival times of sheep with
 natural scrapie.";
 RL J. Gen. Virol. 77:2659-2673(1996).
 CC -I- FUNCTION: The function of PrP is not known. PrP is encoded in the
 host genome and is expressed both in normal and infected cells.
 CC -I- SUBUNIT: PrP has a tendency to aggregate yielding polymers called
 "rods".
 CC -I- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
 CC -I- POLYMORPHISM AT POSITION 171 MAY BE RELATED TO THE
 ALLELES OF SCRAPIE INCUBATION-CONTROL (SIC) GENE IN THIS SPECIES.
 CC -I- DISEASE: FOUND IN HIGH QUANTITY IN THE BRAIN OF HUMANS AND ANIMALS
 INFECTED WITH DEGENERATIVE NEUROLOGICAL DISEASES SUCH AS KURU,
 CREUTZBIELDT-JAKOB DISEASE (CJD), GERRMANN-STRAUSSLER SYNDROME
 (GSS), SCRAPIE, BOVINE SPONGIFORM ENCEPHALOPATHY (BSE),
 CC TRANSMISSIBLE MINK ENCEPHALOPATHY (TME), ETC.
 CC -I- DISEASE: SCRAPIE IS A TRANSMISSIBLE NEURODEGENERATIVE DISORDER OF
 SHEEP AND GOATS. MOST SHEEP THAT CONTRACT THE DISEASE NATURALLY

DIE BETWEEN 24 AND 50 MONTHS OF AGE. THE INCUBATION PERIOD IN SHEEP DEPENDS ON THE STRAIN(S) OF THE INFECTING PATHOGEN, SHEEP AGE AT EXPOSURE, AND THE SHEEP GENOTYPE. THE SURVIVAL TIME IS MAINLY DETERMINED BY A SINGLE GENETIC LOCUS, SIP, WHICH HAS TWO ALLELES, SUSCEPTIBLE (SA) AND RESISTANT (PA). SHORT INCUBATION PERIOD IS CONFERRRED BY THE PARTIALLY DOMINANT SA ALLELE. SCRAPIE CAN BE SPREAD BETWEEN FLOCKMATES, OR IT CAN BE TRANSMITTED FROM AN INFECTED EWEE TO ITS LAMB.

-I- SIMILARITY: Belongs to the prion family.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation. The European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is not modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to licensee@isb-sib.ch).

CC DR EMBL; M31313; AAB97765.1; ..
 CC DR EMBL; D38179; BAR0736.1; ..
 CC DR EMBL; X79912; CRA56283.1; ..
 CC DR PDB; 1G04; 23-JAN-02.
 CC DR PDB; 1M25; 17-JUL-02.
 CC DR InterPro; IPR000817; Prion.
 CC DR pfam; PF00377; prion; 1.
 CC DR pfam; PF03991; Prion_octapep; 5.
 CC DR PRINTS; PR00341; PRION.
 CC DR SMART; SM00157; PRP; 1.
 CC DR PROSITE; PS00291; PRION_1; 1.
 CC DR PROSITE; PS00706; PRION_2; 1.
 CC FT prion; Glycoprotein; GPI-anchor; Repeat; Signal; 3D-structure.
 CC FT SIGNAL 1 24
 CC FT CHAIN 25 256 MAJOR PRION PROTEIN.
 CC FT CARBOHYD 184 184 N-LINKED (GLCNAC. . .) (PROBABLE).
 CC FT CARBOHYD 200 200 N-LINKED (GLCNAC. . .) (PROBABLE).
 CC FT DISULFID 182 217 BY SIMILARITY.
 CC FT DOMAIN 54 95 5 X 8 AA TANDEM REPEATS OF P-H-G-G-Q-W-G-
 CC FT REPEAT 54 62 Q.
 CC FT REPEAT 63 70 1.
 CC FT REPEAT 71 78 2.
 CC FT REPEAT 79 86 3.
 CC FT REPEAT 87 95 4.
 CC FT VARIANT 112 112 5.
 CC FT VARIANT 136 136 M -> T.
 CC FT VARIANT 136 136 A -> V (IN SCRAPIE; SHORT INCUBATION; SIP
 CC SA ALLELE).
 CC FT VARIANT 137 137 M -> T.
 CC FT VARIANT 141 141 L -> F.
 CC FT VARIANT 154 154 R -> H.
 CC FT VARIANT 171 171 R -> Q (LINKED TO SUSCEPTIBILITY TO
 CC SCRAPIE).
 CC FT VARIANT 171 171 R -> H (IN SCRAPIE; LOW INCIDENCE).
 CC FT VARIANT 211 211 R -> Q.
 CC SQ SEQUENCE 256 AA; 27916 MW; 7FFB1A6CCFDBFB8BB CRC64;

Query Match 8 9%; Score 93.5; DB 1; Length 256;

Best Local Similarity 29.2%; Pred. No. 0.5; Matches 6; Mismatches 57; Indels 51; Gaps 12;

Q/ 17 RAPTDVKFPGQQIVGVWYLLRPGRLGVRAKTSRSRSPQRGRQPIKA---RQPQ 72
 D/ :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:
 D/ 26 KRPK---PGCGWNTEG---SRYFGQQ-----SPGGNRP-POGGGMQPH 64

Q/ 73 GRWIAQ---GYPWPPLYNEGGW----AGWLSPRSPRHGNPDPRRSRNIGKVI 123
 D/ :|||:|||:|||:|||:|||:|||:|||:|||:
 D/ 65 GGGWQPHGGWQGP---HGGGWQPHGGGW---GQGGHSQW--NPKSPKTNMKVA 116

Q/ 124 DTLTCGPADLNGYIPVWGAQPLGGVAAALAHGVRAEDGINY 164
 D/ :|||:|||:|||:|||:|||:|||:
 D/ 117 GAAAG-----AVVGG-LGGYMLGSAMSRPLIHFQNDY 148

PRP2_BOVIN STANDARD; PRT; 256 AA.
 ID _PRP2_BOVIN STANDARD;
 AC Q01880;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DB Major prion protein 2 precursor (PrP) (Major scrapie-associated fibril
 DE protein 2).
 OS Bos taurus. (Bovine).
 OC Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Buteraria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovinae; Bovinae; Bos.
 OX NCBI_TAXID:9913;
 RN [1]
 RP SOURCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=93118243; PubMed=1362024;
 RA Shinagawa M.; Sequence analysis and expression of bovine PrP gene in
 RA mouse L-929 cells. ";
 RL Virus Genes 6:343-356(1992).
 CC -!- FUNCTION: The function of PrP is not known. PrP is encoded in the
 CC host genome and is expressed both in normal and infected cells.
 CC -!- SUBUNIT: PrP has a tendency to aggregate yielding polymers called
 CC "rods".
 CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
 CC -!- DISEASE: FOUND IN HIGH QUANTITY IN THE BRAIN OF HUMANS AND ANIMALS
 CC INFECTED WITH DEGENERATIVE NEUROLOGICAL DISEASES SUCH AS KURU,
 CC CERBUTZELDT-JAKOB DISEASE (CJD), GERSTMANN-STRAUSSLER SYNDROME
 CC (GSS), SCRAPE, BOVINE SPONGIFORM ENCEPHALOPATHY (BSE),
 CC TRANSMISSIBLE MINK ENCEPHALOPATHY (TME), ETC.
 CC -!- SIMILARITY: Belongs to the prion family.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC RESULT 23
 PRIO_BOVIN STANDARD; PRT; 264 AA.
 ID PRIO_BOVIN STANDARD;
 AC P10279;
 DT 01-MAR-1989 (Rel. 10, Created)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DB Major prion protein 1 precursor (PrP) (Major scrapie-associated fibril
 PRNP OR PRP).
 OS Bobtaurus (Bovine).
 OC Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Buteraria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovinae; Bovinae; Bos.
 RN [1]
 RP SOURCE FROM N.A.
 RC STRAIN-Holstein-Friesian;
 RX MEDLINE=9116314; PubMed=1671225;
 RA Goldmann W., Hunter N., Martin T., Dawson M., Hope J.;
 RT "Different forms of the bovine PrP gene have five or six copies of a
 short, G-C-rich element within the protein-coding exon.;"
 RL J. Gen. Virol. 72:201-204(1991).
 RN [2]
 RP SOURCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=93118243; PubMed=1362024;
 RA Shinagawa M.;
 RA Yoshiimoto J., Iinuma T., Ishiguro N., Horiuchi M., Imamura M.,
 RA "Comparative sequence analysis and expression of bovine PrP gene in
 mouse L-929 cells.";
 RL Virus Genes 6:343-356(1992).
 RN [3]
 RP SOURCE FROM N.A.
 RX MEDLINE=93179783; PubMed=8440932;
 RA Prusiner S.B., Ruzi M., Scott M., Serban D., Serban H.,
 RA Taraboulos A., Gabriel J.M., Wells G.A., Wilsmith J.W., Bradley R.;
 RT "Immunological and molecular biological studies of prion proteins in
 bovine spongiform encephalopathy.;"
 RL J. Infect. Dis. 167:602-613(1993).
 RN [4]
 RP SOURCE FROM N.A.
 RC STRAIN=Holstein-Friesian; TISSUE=Brain;
 RC STRAIN=Holstein-Friesian; TISSUE=Brain;
 RI Horiuchi M.;
 RI Submitted (FEB-1997) to the EMBL/GenBank/DDBJ databases.
 RN [5]
 RP SOURCE FROM N.A.
 RC STRAIN=Jersey;
 RX MEDLINE=21422903; PubMed=11531705;
 RA Hills D., Comincini S., Schlaepfer J., Dolf G., Ferretti L.,
 RA Williams J.L.;
 RT "Complete genomic sequence of the bovine prion gene (PRNP) and
 RT polymorphism in its promoter region.;"
 RL Anim. Genet. 32:231-232(2001).
 RN [6]
 RP SOURCE FROM N.A.
 RC STRAIN=Korean;
 RA Yoo H.S., Kang S.G., Choi I.S., Kang S.K., Huang W.S.;
 RA "Nucleotide Sequence of PrP cDNA in Korean cattle";
 RL Submitted (JUN-2002) to the EMBL/GenBank/DDBJ databases.

Query Match 8.9%; Score 93.5; DB 1; Length 256;
 Best Local Similarity 29.2%; Pred. No. 0; Mismatches 6; Indels 51; Gaps 12;

Matches 47; Conservative 6; Mismatches 57; Indels 51; Gaps 12;

RN [7]
 RP SEQUENCE OF 1-15 FROM N.A.
 RA Tanaka M., Inoue S., Ikeda T., Horiuchi M., Ishiguro N., Shinagawa M.;
 RL Submitted (JAN-1994) to the EMBL/GenBank/DBJ databases.
 RN [8]
 SEQUENCE OF 25-36.
 RX MEDLINE=89057122; PubMed=2904126;
 RA Hope J., Reekie L.J.D., Hunter N., Multhaup G., Beyreuther K.,
 White H., Scott A.C., Stack M.J., Dawson M., Wells G.A.;
 RT "Fibrils from brains of cows with new cattle disease contain scrapie-associated protein.";
 RT Nature 336:390-392(1988).
 RN [9]
 RP STRUCTURE BY NMR OF 132-241.
 RX MEDLINE=20359707; PubMed=10899999;
 RA Lopez Garcia F., Zahn R., Riek R., Wuthrich K.;
 RT "NMR structure of the bovine prion protein.";
 RL Proc. Natl. Acad. Sci. U.S.A. 97:8314-8319(2000).
 CC -!- FUNCTION: The function of PrP is not known. PrP is encoded in the host genome and is expressed both in normal and infected cells.
 CC -!- SUBUNIT: PrP has a tendency to aggregate yielding polymers called "rods".
 CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
 CC -!- DISEASE: FOUND IN HIGH QUANTITY IN THE BRAIN OF HUMANS AND ANIMALS INFECTED WITH DEGENERATIVE NEUROLOGICAL DISEASES SUCH AS KURU (GSS), SCRAPE, BOVINE SPONGIFORM ENCEPHALOPATHY (BSE), TRANSMISSIBLE MINK ENCEPHALOPATHY (TME), ETC.
 CC -!- SIMILARITY: Belongs to the prion family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC
 DR EMBL; X55802; CAA39368.1; -.
 DR EMBL; D10612; BAA01467.1; -.
 DR EMBL; D10613; BAA01468.1; -.
 DR EMBL; S55629; AAB25524.1; -.
 DR EMBL; AB001468; BAA19253.1; -.
 DR EMBL; AJ298978; CAC37367.1; -.
 DR EMBL; AF517842; AAM62709.1; -.
 DR EMBL; D26151; BAA05138.1; -.
 DR PIR; A54330; A54330.
 DR PDB; 1DWZ; 26-FEB-02.
 DR PDB; 1DX0; 26-FEB-02.
 DR PDB; 1DX1; 26-FEB-02.
 DR Interpro; IPR000817; Prion.
 DR Pfam; PF00377; Prion_1.
 DR Pfam; PF03991; Prion_octapep; 6.
 DR Prints; PRO0341; Prion.
 DR SMART; SM00157; PrP_1.
 DR PROSITE; PS00291; PRION_1; 1.
 DR PROSITE; PS00706; PRION_2; 1.
 DR DOMAIN
 KW P104; Glycoprotein; GPI-anchor; Repeat; Signal; Polymorphism; 3D-structure.
 FT SIGNAL 1 24
 FT CHAIN 25 264
 FT CARBOHYD 192 192
 FT CARBOHYD 208 208
 FT DISULFID 190 225
 FT DOMAIN 54 103
 PT REPEAT 54 62
 PT REPEAT 63 70
 PT REPEAT 71 78
 PT REPEAT 79 86
 PT REPEAT 87 94
 PT REPEAT 95 103
 PT VARIANT 71 78
 PT CONFLICT 218 218
 PT HELIX 136 138
 PT STRAND 140 141
 PT HELIX 155 162
 PT TURN 163 164
 PT HELIX 165 167
 PT STRAND 173 174
 PT HELIX 184 203
 PT TURN 204 206
 PT HELIX 211 237
 SQ SEQUENCE 264 AA; 28614 MW; D6D214038316A231 CRC64;
 RN
 RP STRUCTURE BY NMR OF 132-241.
 RX MEDLINE=20359707; PubMed=10899999;
 RA Lopez Garcia F., Zahn R., Riek R., Wuthrich K.;
 RT "NMR structure of the bovine prion protein.";
 RL Proc. Natl. Acad. Sci. U.S.A. 97:8314-8319(2000).
 CC -!- FUNCTION: The function of PrP is not known. PrP is encoded in the host genome and is expressed both in normal and infected cells.
 CC -!- SUBUNIT: PrP has a tendency to aggregate yielding polymers called "rods".
 CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
 CC -!- DISEASE: FOUND IN HIGH QUANTITY IN THE BRAIN OF HUMANS AND ANIMALS INFECTED WITH DEGENERATIVE NEUROLOGICAL DISEASES SUCH AS KURU (GSS), SCRAPE, BOVINE SPONGIFORM ENCEPHALOPATHY (BSE), TRANSMISSIBLE MINK ENCEPHALOPATHY (TME), ETC.
 CC -!- SIMILARITY: Belongs to the prion family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC
 DR EMBL; X74771; CAA52781.1; -.
 DR S37137; S37137.
 DR HSSP; P10279; 1DWY.
 DR Interpro; IPR000817; Prion.

| | | | |
|--|--|--|--|
| Db | 110 TNMKHVAGAAAG-----AVVGG-LGGYMLGSAMRPLIHFNDY 148 | Db | 26 KRDK---PGGGINNTGG---SRYPGQG---SPGNNRYP-POGGGGWQPH 64 |
| RESULT 28 | | Qy | 73 GRWIAQPYPPWPLIXGNBCCG----AGWLISPRGRSPH----WGP-----NDPBRRS 116 |
| PRI_OPOHE | STANDARD; | ID | 65 GGGWQP-----HGGWSEPHGSGW-----GPHGGCGWQSGTHSQWNPKSPK 109 |
| ID PRI_OPOHE | PRT; 256 AA. | AC | P47852; |
| DT 01-FEB-1996 (Rel. 33, Created) | | DT 01-FEB-1995 (Rel. 33, Last sequence update) | 117 RNIGKVITLTCGFADILGYIPVGAPIGGVAALAHGVRAEDGINY 164 |
| DT 15-MAR-2004 (Rel. 43, Last annotation update) | | DB 110 TNMKHVAGAAAG-----AVVGG-LGGYMLGSAMRPLIHFNDY 148 | |
| DE Major prion protein precursor (PrP). | | GN odocoileus hemionus (Mule deer) (Black-tailed deer). | |
| OS | | OC Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi; | |
| OC | | OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Cervoidea; | |
| RT | | RT Cervidae; Odocoileinae; Odocoileus. | |
| OX NCBI_TaxID=9872; | | RL [1] | |
| RR | | SEQUENCE FROM N.A. | |
| RP MEDLINS:97333774; PubMed:9250209; | | RP P04233; | |
| PI Cervenakova L., Rohwer R., Williams S., Brown P., Gajdusek D.C.; | | AC 01-FEB-1995 (Rel. 31, Created) | |
| OS "High sequence homology of the PrP gene in mule deer and Rocky | | DT 01-FEB-1995 (Rel. 31, Last sequence update) | |
| RT Mountain elk.;" | | DT 15-MAR-2004 (Rel. 43, Last annotation update) | |
| RL Lancet 350:219-220(1997). | | DE Major prion protein precursor (PrP) (Major scrapie-associated fibril protein 2). | |
| CC -!- FUNCTION: The function of PrP is not known. PrP is encoded in the host genome and is expressed both in normal and infected cells. | | CC DE Major prion protein 2 precursor (PrP) (Major scrapie-associated fibril protein 2). | |
| CC -!- SUBCELLULAR LOCATION: PrP has a tendency to aggregate yielding polymers called "rods". | | CC OC Bovidae; Bovinae; Tragelaphus. | |
| CC -!- SIMILARITY: Belongs to the prion family. | | CC OX NCBI_TaxID=9946; | |
| CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch). | | CC RN [1] | |
| CC DR SMART; SMO157; PRP; 1. | | RC TISSUE:Brain; | |
| DR EMBL; U25965; AAH68941; 1. | | RA Martin T.C., Hughes S.L., Hughes K.J., Dawson M.; | |
| DR PROSITE; PS00706; PRION_2; 1. | | RL Submitted (AUG-1993) to the EMBL/GenBank/DBJ databases. | |
| DR KW Prion; Glycoprotein; GPI-anchor; Repeat; Signal. | | CC -!- FUNCTION: The function of PrP is not known. PrP is encoded in the host genome and is expressed both in normal and infected cells. | |
| FT SIGNAL 1 BY SIMILARITY. | | CC -!- SUBUNIT: PrP has a tendency to aggregate yielding polymers called "rods". | |
| FT CHAIN 25 MAJOR PRION PROTEIN. | | CC -!- SIMILARITY: Attached to the membrane by a GPI-anchor. | |
| FT CARBOHYD 184 N-LINKED (GLCNAC. . .) (POTENTIAL). | | CC -!- DISEASE: FOUND IN HIGH QUANTITY IN THE BRAIN OF ANIMALS INFECTED WITH DEGENERATIVE NEUROLOGICAL DISEASES SUCH AS KURU, CREUTZFELDT-JAKOB DISEASE (CJD), GERSMANN-STRAUSSLER SYNDROME (GSS), SCRAPE, BOVINE SPONGIFORM ENCEPHALOPATHY (BSE), TRANSMISSIBLE MINK ENCEPHALOPATHY (TME), ETC. | |
| FT DISULFID 182 BY SIMILARITY. | | CC -!- SIMILARITY: Belongs to the prion family. | |
| FT DOMAIN 54 5 X 8 AA TANDEM REPEATS OF P-H-G-G-G-W-G- | | CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch). | |
| FT REPEAT 54 1. | | CC DR EMBL; X74759; CAAS52775; 1. | |
| FT REPEAT 63 2. | | DR HSSP; P10279; IDMY. | |
| FT REPEAT 71 3. | | DR InterPro; IPR000817; Prion. | |
| FT REPEAT 79 4. | | DR PROSITE; PS00706; PRION_2; 1. | |
| FT REPEAT 87 5. | | DR PRFam; PR03991; Prion; 1. | |
| SQ SEQUENCE 256 AA: 27961 MW: E98EB121C92FD36 CRC64; | | DR PRFam; PR03991; Prion; octapep; 5. | |
| Query Match 8.7%; Score 90.5; DB 1; Length 256; | | DR PRINTS; PR03341; PRION. | |
| Best Local Similarity 27.4%; Pred. No. 0.89; Mismatches 51; Indels 65; Gaps 12; | | DR SMART; SMO157; PRP; 1. | |
| Matches 46; Conservative 6; Mismatches 51; Indels 65; Gaps 12; | | DR PROSITE; PS00706; PRION_2; 1. | |
| OY 17 RPRTDVKEPGGQQTIVGGVYLPRKGPRGVRATRKTSERSQPRQRQDTPKA---RQPQ 72 | | DR PRFam; PR03991; Prion; GPI-anchor; Repeat; Signal.. | |
| FT REPEAT 54 1. | | DR SIGNAL 1 MAJOR PRION PROTEIN 2. | |
| FT REPEAT 63 2. | | FT CHAIN 25 N-LINKED (GLCNAC. . .) (POTENTIAL). | |
| FT REPEAT 71 3. | | FT CARBOHYD 184 N-LINKED (GLCNAC. . .) (POTENTIAL). | |
| FT REPEAT 79 4. | | FT CARBOHYD 200 5 X 8 AA TANDEM REPEATS OF P-H-G-G-G-W-G- | |
| FT REPEAT 87 5. | | FT DISULFID 182 217 | |
| SQ SEQUENCE 256 AA: 27961 MW: E98EB121C92FD36 CRC64; | | FT DOMAIN 54 95 O. | |

Query Match 8.7%; Score 90.5; DB 1; Length 256;
 Best Local Similarity 28.1%; Pred. No. 0 89; Mismatches 59; Indels 49; Gaps 11;

Matches 45; Conservative 7; Job time : 20 secs
 Job time : 20 secs

Qy 17 RRPDTDKRPGGSDIIVGGVYLLPRGRPRIGYVRAIRKTTSRSQRGRROPIKA---RQOG 73
 Db 26 KRPK---PGGNNNTGG---SRYPGOG-----SPGPNRYPQEGDWGQPHG 65
 Qy 74 RIIWAQP---GYPPPLYNGECCW---AGNLSPRCSRPRWGPNPRRSRNLKVID 124
 Db 66 GWGOPRIVGGWQP---HGGWQPHRGGGW--GCGGTHGW--NKEPSKPCTNMKGIVAG 117
 Qy 125 TTTTCGFAIDLIMGYTPVVGAPLGLGVAALAHGVRAIEDGINY 164
 Db 118 AAAG----AVVGG-LGGIMLGSAMSRSRPUHFQSDY 148

RESULT 30

V 770K TYMV STANDARD; PRT; 628 AA.

A p10357; P10357;

DT 01-MAR-1989 (Rel. 10, Created)

DT 01-AUG-1992 (Rel. 23, Last sequence update)

DB 69 kDa protein.

OS Turnip yellow mosaic virus.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tymoviridae;

OC Tymovirus.

OX NCBI_TaxID:12154;

RN [1]

RP SEQUENCE FROM N.A.

RX PMID:88289359; PubMed=3399388;

RA March M.D.; Boyer J.C.; Haenni A.L.;

RT "overlapping open reading frames revealed by complete nucleotide sequencing of turnip yellow mosaic virus genomic RNA.";

RL Nucleic Acids Res. 16:6157-6173 (1988).

CC -- FUNCTION: Not known.

CC -- SIMILARITY: TO 65 TO 70 kDa PROTEIN FROM OTHER TYMOWIRUSES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC

CMBL: X07441; CAA30321.1; ALT_SEQ.

DR: PIR: S01955; S01955.

DR: InterPro: IPR004335; Tympo_45_70kDa.

DR: Pfam: PF03551; Tympo_45_70kDa; 1.

SQ SEQUENCE 628 AA; 69195 MW; 9B01CE55ADFCBAC77 CRC64;

Query Match 8.7%; Score 90.5; DB 1; Length 628;
 Best Local Similarity 25.7%; Pred. No. 2.2; Mismatches 55; Indels 69; Gaps 9;
 Matches 48; Conservative 15; Mismatches 55; Indels 69; Gaps 9;

Qy 4 LPKPKQKTKRGNTRRPTDKVPRGG---GIVGGVYL-----P 38

Db 285 LPPNRRHRTSTGHIPPTTSRPTGPPSSRLQPRVHYOSSPRHTNFRPSSIRKDALLQTGP 344

Qy 39 RRG---PRIGVRAKTRKTSRSQRGR-----OPIPKAR-QPGRHWAQPGYFWP 84

Db 345 RLGHLERLGQPAULRTSERSPPTKRRFLRFSRSENRPLRPLPEATLAPSRYH---RRPYP 400

Qy 85 IYKNEGCCWAGWILS-----PROSRPRHGPNDPDRR-----SRNUGV 123

Db 401 LPNPPAALPSIAYTSSRKHHSLPKGAKPKEGAPPFRRLPSPAPRQLPLDGR-- 458

Qy 124 DTUTCGF 130.

Db 459 ---TPGF 462